MODELLING GROWTH AND YIELD OF DOUGLAS-FIR USING DIFFERENT INTERVAL LENGTHS IN THE SOUTH ISLAND OF NEW ZEALAND

A thesis

submitted in partial fulfilment

of the requirement for the degree

of

Doctor of Philosophy in forestry

in the

University of Canterbury

by

Sang Hyun Lee

University of Canterbury

1998

Abstract

This study describes several refinements and improvements in whole stand growth and yield modelling of Douglas-fir grown in four regions of the South Island of New Zealand, namely Canterbury, Nelson, Southland and Westland. Modelling growth and forecasting yields are necessary for providing adequate tools with which to manage wood production from forests.

The study comprised three major components: 1) development of whole stand growth and yield models with data sets of various interval lengths; 2) cross fitting models with different data sets reciprocally; and 3) check estimates using a growth and yield model derived from a data set free of auto-correlation.

The methodology emphasised in developing the equations in this study involved re-arrangement of the data to reflect different interval lengths among re-measurements for modelling purposes. Modification of data sets allowed an investigation into which growth intervals should be used to obtain the least biased models overall, and efficiently. The approach involved fitting single equations to each of three state variables, mean top height (h_{100}), basal area/ha (G) and stocking/ha (N). Differences in growth trajectories across the four regions were identified and incorporated into single variable equations using dummy variables for improving the fitting of mean top height (h_{100}), basal area/ha (G), and stocking/ha (N) equations. The main finding from this study was the level of improvement in making predictions through adoption of a mixed interval projection equation strategy compared with other options.

Examining consistency among the predicting equations which had been developed from the different interval data sets, involved testing each form of model individually

for all the data sets. The models based on mixed intervals were found to fit well for all the other interval length data sets. A subset of uncorrelated data was then created by selecting one re-measurement from each permanent sample plot (PSP), which was then used to validate the appropriateness of the equations derived from the full data sets, in order to overcome problems of dealing with correlated data. Coefficients for each of the equations for mean top height, basal area/ha and stocking/ha which were derived from this check data set were found to be very similar to the regression coefficients obtained from the full data set.

Although the growth models developed in this study may require further examination, they do provide a very useful guide for selecting appropriate re-measurement interval lengths to derive satisfactory models which are the least biased overall. It is strongly recommended that modellers in the future adopt a mixed interval strategy as one data set option to evaluate.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

This study deals with refinements in modelling the growth and yield of Douglas-fir throughout the South Island of New Zealand. It builds on the findings of Temu (1992) who also developed a South Island model which was an improvement on the previous model of the New Zealand Forest Research Institute (NZFRI) (Law, 1990).

Douglas-fir (Pseudotsuga menziesii (Mirbel) Franco) ranks as the second most important plantation tree species after radiata pine in New Zealand. Although radiata pine plantations dominate the New Zealand forestry landscape, the area of Douglas-fir plantations is around 70,000 hectares, or 4 percent of the total plantation forest area. More than half the area, 37,093 hectares is established in the South Island, with 32,873 hectares located in the North Island (Neilson and BMS Ltd., 1994).

The first recorded introduction of Douglas-fir into New Zealand was of a single plant, shipped in 1859 from Veitch's nursery near Exeter, England to J.B.A Acland of Mt Peel station, Canterbury. From about 1870 onwards Douglas-fir was used for amenity and farm plantings in the Southland, especially in the Canterbury region where old trees and plantations may still be seen in the foothills (Miller and

Knowles, 1994). Before 1926 no accurate records were kept of the origins of seed reaching New Zealand. The main importer of seed in this early period was the Lands and Survey Department. It is commonly believed that most Douglas-fir seed obtained by the Department between 1902 and 1928 came from Washington and Oregon, with a lesser amount from British Columbia (Wilcox, 1978; James and Bunn, 1978; Miller and Knowles, 1994). Between 1926 and 1974, a total of 3000 kg of seed was imported by the New Zealand Forest service from North America. Approximately 97 percent of this imported seed was from sources in Washington and Oregon, thereby forming the basis of the New Zealand Douglas-fir industry. Of the remainder, 1 percent was from British Columbia and 2 percent from California (Miller and Knowles, 1994).

Research investigating performance of Douglas-fir provenances in New Zealand was started in 1955 by the NZFRI. Three trial series were established and examined:

1) first trial series (1957) – 35 provenances constituted mostly of commercial seedlots from Washington and Oregon including 4 New Zealand land races;

2) second trial series (1959) – consisted 45 provenances from California and Coastal Oregon and 1 New Zealand land races;

3) third trial series (1971 and 1974) – to compare the qualities of local seed sources in New Zealand.

The assessment showed that when the 1957 trials were assessed at age 6 years, the Washington seedlots were generally the tallest, but the 4 New Zealand land races were among the best performers on most sites. In the 1959 trials, provenances from low elevation localities in the California coastal fog-belt performed well at all sites. The single New Zealand land race represented in the trials (from Kaingaroa, and presumed to be of Washington ancestry) fared well in aggregate performance. The results of the third trial showed that seedlots from a number of South Island sources (e.g. Queenstown, Naseby and Wanaka) were slow growing. Although not included in the third trial tests, the few seed stands of low-elevation coastal California origin

available in New Zealand were considered to be valuable, their parent seedlots having performed well in provenance trials (Miller and Knowles, 1994).

Douglas-fir is more tolerant of cold winters and tends to be more wind stable than radiata pine on exposed sites. On some sites, usually at higher altitudes which are sub-optimal for radiata pine because of risks of wind and snow damage, Douglas-fir can be a better performer. Significant areas of farmland suitable for Douglas-fir plantations are potentially available in the South Island (Neilson and BMS Ltd., 1994).

Douglas-fir is considered to have good timber qualities, especially for engineering purposes (Hellawell, 1978) or as least as good as those of radiata pine (James and Bunn, 1978). However, there are two major reasons why radiata pine is the dominant species, making up 91 percent of the planted forest area. One is that Douglas-fir grows more slowly than radiata pine; technical rotations are 40-60 years, whereas those for radiata pine are about 30 years. Another is that Douglas-fir was infected with Phaeocryptopus gaeumannii, Swiss needle-cast fungus, in the 1960's which led to significant reductions in growth of the species. Comparisons made before and after the onset of Swiss needle-cast fungus, indicated that basal area increment of Douglas-fir had been lowered by at least 25 percent (Hood and Kershaw, 1975; James and Bunn, 1978; Beekhusis, 1978; Miller and Knowles, 1994).

The best information on the effect of Swiss needle-cast is in Liu Xu's (1990) work. Liu Xu developed separate models for the growth and yield in diseased and undiseased conditions. The basis of separating data from diseased and undiseased stands was the time that the infection occurred (e.g. pre-1963 and post-1963). Yields forecasted using the diseased and undiseased components of Liu Xu's model showed some indication of the loss of productivity resulting from the presence of Swiss needle-cast fungus. Comparison of the yield differences between diseased and undiseased stands using some specified stand conditions (e.g. initial age of 20 years, site index of 30 m, initial stocking 1500 stems/ha and initial basal area of 45 m^2/ha), showed that the total volume production of a young diseased stand (30 years of age) was 17% less than that of the same undiseased stand, while the net volume yield was

about 23% less. Total volume production of old diseased stands (60 years of age), was 33% less, while net volume yield was about 39% lower.

However, Douglas-fir has an excellent reputation as a general purpose structural timber in both domestic and export markets. For many years most Douglas-fir harvested was converted into framing timber for New Zealand's domestic and Australian markets. Douglas-fir exports from New Zealand include both sawn timber and logs. The three largest markets for New Zealand exports of sawn timber from 1991 to 1993 were Australia, New Caledonia and Japan respectively, representing an annual export volume of almost 70,000 m^3. New Zealand Douglas-fir has a great export potential to the Japanese market, because the falling supply of North American Douglas-fir has encouraged the Japanese sawmilling industry to look for new sources (Miller and Knowles, 1994).

There are three major models available for the prediction of the growth of Douglas-fir: The North Island model (Liu Xu, 1990); the South Island model (Law, 1990); and the Temu model (Temu, 1992). An investigation examining the accuracy of these three models by Tennent (1995) showed that all three models predicted growth to within generally acceptable limits. The models performed best in the regions that their original data were derived form. However the prediction of Law's model was poor and did not distinguish regional growth trends.

These three models used only nonoverlapping growth intervals in the parameter estimation phase of model development as the most common procedure for estimating parameters in algebraic difference equations. Another possibility is to use all possible growth intervals. If permanent sample plots have been measured n times, there are $\binom{2}{n}$ combinations of different time intervals between T_i and T_{i+1} that can be derived and used to build equations.

The use of all possible combinations of data intervals was found a powerful approach in previous studies (Villanueva, 1992 and Chikono, 1994). However, what has not

been obvious is which growth intervals should be used to obtain the parameter estimates.

As already mentioned above the South Island has a great potential to grow Douglas-fir. Therefore this study, was aimed at producing a model of growth and yield of South Island Douglas-fir crops, that improved on existing tools by using different growth interval data sets in order to examine which measurement interval would be appropriate to build more precise models.

1.2 Objectives

The overall aim of this study was to produce biologically and mathematically sound and sufficiently reliable stand growth and yield models, using different interval lengths based on re-measured growth data for Douglas-fir grown in the South Island of New Zealand. The specific objectives of this study were to:

1. provide a strategy for selecting appropriate projection intervals between initial and subsequent ages of measurements, with which to analyse differences in mean top height (h_{100}), net basal area/ha (G) and stocking/ha (N);

2. evaluate gains in precision and lack of bias when modelling growth with various projection intervals based on analyses of a data set in New Zealand;

3. examine consistency among the predicting equations which were developed from the various interval length data sets.

1.3 The Scope of the Study

Populations for the whole stand growth and yield modelling conducted in this study refer to Douglas-fir grown throughout the South Island of New Zealand. Permanent Sample Plot (PSP) data used came from four regions located in the South Island:

Canterbury, Nelson, Southland and Westland. Measurements are available for ages ranging from 5 to 78 years and re-measurement intervals ranged from 1 to 28 years.

1.4 Notation

Symbols and definitions used throughout in this study are adopted from standard IUFRO notation. Unless otherwise stated, the following symbols and definitions apply.

α_i, β_i, γ_i : regression coefficients;

AL : altitude above sea level (m)

SEE : standard error of parameter estimates;

ESS : error of sum of squares;

MSE : mean square error;

f : form factor;

G_i : net basal area per hectare at crop age T_i (m^2/ha);

$h_{(100,i)}$: mean top height at crop age T_i (m);

N : total number of observations in a population or stocking per hectare;

N_i : number of stems per hectare (stocking) at crop age T_i;

NID (0, δ^2) : normally and independently distributed with mean 0 and constant variance δ^2;

RND : random normal deviate;

S : site index –mean top height at age 40 years;

T_i : age of crop in years for period i;

T_t : age of thinning in years;

V_i : stand volume per hectare at crop age T_i (m^3/ha)

Chapter 2

Review of Literature

2.1 Forest Growth and Yield Modelling

A model may be defined as any abstraction or simplification of some aspects of a real system. "Growth model" generally refers to a system of equations that can predict the growth and yield of a forest stand under a wide variety of conditions, and may comprise a series of mathematical equations. A stand growth model is an abstraction of natural dynamics of a forest stand, and may encompass growth, mortality and other changes in stand composition and structure (Hall and Day, 1977; Vanclay, 1994). Growth models can be used to make predictions, formulate predictions and guide forest policy.

The purpose of growth and yield models can be regarded as, given set of stand or tree characteristics, such as basal area and stems per hectare, at one point in time (T_1) and to certain locality, to determine by how much these will have changed at time T_2 in the future given specified stand measurements and treatments. Such a quantitative understanding has been crucial to forest managers, in helping them make important forest management decisions.

Effective decision-making for forest management may depend on reliable forecasts of growth and yield. Thus, modelling growth and yield forecasting is necessary for exploring silvicultural options, harvest planning and updating inventory information. For example, several tending regimes can be simulated and their results compared to objectively determine the most appropriate silvicultural practice, while forest managers can easily forecast the nature and timing of future harvests with growth and harvest models. Growth models can also be used to update management planning inventory data between two successive inventories.

2.2 Growth Modelling Approach

2.2.1 Types of Growth Models

There are many growth models, and it is useful to classify models by the level of detail they provide.

2.2.1.1 Process and Empirical Models

Bruce and Wensel (1987) suggested process and empirical models (Draper and Smith, 1981), which were also referred to as functional and predictive models, and process and prediction models (Vanclay, 1994).

Process models simulate the biological process of growth using inputs such as: temperature and soil nutrient levels, photosynthesis rate, as well as respiration and allocation of photosynthesis to roots and stems (Landsberg, 1986; Belli and Nautiyal, 1989; Bossel, 1991; Vanclay, 1994). The main aim of process models is understanding and explaining the behaviour of particular ecosystems, with less emphasis on predicting quantitative outcomes. They do however supply an acceptable degree of accuracy for certain forest management decisions (Mohren and Burkhart, 1994). These models

provide a better understanding of growth and stand dynamics, but have not yet successfully been used for predicting timber yields for forest management.

Empirical models are based on periodic tree measurements, and make no attempt to represent every factor that may affect tree growth. These models may sacrifice details of growth processes to acquire greater efficiency and accuracy in providing information for forest management (Vanclay, 1994). The power of a conventional empirically based model is that is draws on historical yield data to introduce site specific determinants of yield that are hard to model with process-based models. The different emphases contained within these two varying models and the fact that each of them is useful for its own purpose, makes it seem unlikely that one will replace the other (Goulding, 1994; Mohern and Burkhart, 1994; Adlard, 1995).

A hybrid model combines the best features of both models, by using simple processes to generate biological realism for use as input into an empirical model (Belli and Nautiyal, 1989). Hybrid models provide an increase in biological reality over traditional empirical yield models and may allow more precise estimations of yield because of their greater sensitivity to temporal variation in growth conditions. For example, Woollons et al. (1998) showed improvement in precision when modelling basal area/ha, by including temperature, solar radiation and rainfall of radiata pine grown in the Nelson region of the South Island, New Zealand.

2.2.1.2 Deterministic and Stochastic Models

A deterministic model predicts expected values under a given set of conditions. Given the same initial conditions, a deterministic model will always predict the same result. However because of natural variation in the environment, such as fluctuations in temperature or rainfall, real forest stands may not grow at an exactly uniform rate.

A stochastic model attempts to illustrate natural variation by providing different predictions, each with a specific probability of occurrence. Any one of these estimates may correspond to the growth under some circumstances, but they may differ from the expected growth. Although stochastic models can provide some useful information not available from deterministic models, most of the information needs for forest managers can be provided more efficiently with deterministic models.

2.2.1.3 Static and Dynamic Models

Static growth models attempt to predict directly over the course of time the quantities of interest (volumes, mean diameter). This approach can give good results for unthinned stands, or for stands subject to a limited range of standardised treatments for which long-term experimental data are available.

Dynamic models forecast stand growth over a wider range of tending regimes (initial spacing, various thinning and pruning sequences and intensities). The trajectories over time are then obtained by adding or integrating these rates (Garcia, 1988).

2.3 Growth Modelling Alternatives

Growth models can be classified according to the level of details provided. Munro (1974) and Daniels and Burkhart (1988), categorised growth and yield models into three types.

1) Whole stand models
2) Diameter distribution models
3) Individual tree growth models
   - individual tree distance dependent models
   - individual tree distance independent models

These models are discussed in more detail in the following sections.

2.3.1 Whole Stand Growth Models

Stand growth models describe the development of stand level measurements such as basal area/ha and stems/ha, by equations which use stand level variables such as age, mean top height and basal area/ha to predict future stand level values. These models require relatively little information to simulate growth of a stand, and consequently provide rather general information about the future of the stand.

Whole stand models are the most widely represented kind of growth model and are appropriate for management planning of forest plantations (Garcia, 1988). The advantages of whole stand modelling are generally simplicity, robustness and the ability to use conventional inventory information, while a disadvantage is that they do not provide reliable information at the individual tree level.

2.3.1.1 Yield Tables

Early attempts to model stand productivity were tried using yield tables, which present the expected yields from even-aged stands at various ages, and is the oldest method of yield prediction.

2.3.1.1.1 Normal Yield Tables

The normal yield tables were published in Germany in 1787 (Bickford et al., 1957 Curtis, 1972; Clutter et al., 1983; Avery and Burkhart, 1994), which provide estimates of expected yields tabulated by stand age and site index obtainable from fully stocked or normal forest stands. Normal yield tables were generally constructed by graphical techniques using data from temporary plots (Vanclay, 1994).

The application of the notion of normality in European forest management was a result of inadequate wood supply (Whyte and Woollons, 1990). Ware et al. (1988) however, pointed out difficulties in the application of normal yield tables:

1) this approach was restricted to fully stocked or normally stocked stands, adjustment is needed when the tables were used for non-normal stands;

2) yields were analysed using graphical techniques, which constrains the number of variables used;

3) assessment of normality or full stocking was subjective, and did not represent a rational management goal, because a non-fully stocked stand with proper silviculture treatment could provide wider reforms than a fully stocked untended stand (Curtis, 1972).

A similar approach used empirical yields tables (Schumacher, 1939; Bennet et al., 1959; Avery and Burkhart, 1994). An advantage of empirical yield tables over normal yield tables was that they attempted to characterise the expected yield from stands with an average stocking rather than fully stocked stands. This approach was more objective than that used for normal yield table, but the resulting tables were insensitive to variation in stand density, and adjustment was required when the tables were applied to stands that did not use the average stocking.

2.3.1.1.2 Variable Density Yield Tables

The notion of normal stocking in normal yield tables was found to be erroneous by many foresters and attempts were made to clarify the situation. The first variable density yield prediction equations were developed by MacKinney et al. (1937) and applied by MacKinney and Chaiken (1939), and Schumacher (1939) by liner regression. Their equation was based on a hypothesis that relative growth rate varies inversely with age.

$$\frac{\partial V}{V} \propto T^{-2} \tag{2.1}$$

Thus,

$$\text{In } V = \beta_0 - \beta_1 T^{-1} \tag{2.2}$$

where,

V= volume (m^3)

T= age (years)

The parameter β_0 defines the upper asymptote which, in this case, is volume at the end of rotation and β_1 determines the rate of growth. For example, an equation used by MacKinney and Chaiken (1939) was

$$In(Y) = \beta_0 + \frac{\beta_1}{T} + \beta_2 S + \beta_3 \log(SDI) + \beta_4 C \tag{2.3}$$

Where Y is yield, T is stand age, S is site index, SDI is Reineke's (1933) stand density index and C is composition index. In this approach, stand density was a dynamic part of the prediction system and multiple regression analysis was used to fit the equations and estimate parameters. This equation included other desirable properties apart from expressing density as an independent variable: 1) the dependent variable yield (Y) was predicted from a specific combination of independent variables over a wide range; 2) the logarithm of yield was proportional to the reciprocal of age; 3) the functions exhibited asymptotic growth.

Many researchers have used similar methods for constructing growth and yield equations since the initial proposal by Schumacher (1939) (for example, Schumacher and Coile, 1960; Bailey and Ware, 1983; Murphy and Farrar, 1988; Avery and Burkhart, 1994).

2.3.1.2 Compatible Growth and Yield Models

The effective expression of the biological relationship between growth and yield by mathematical relations between differentiation and integration, was suggested by Buckman (1962) and Clutter (1963). The illustration of compatibility between growth and yield, is that yield is related to growth and is the accumulated growth up to a specified point in time, and can be derived mathematically by integrating a growth function. The derivative of a yield function, therefore gives a growth function (Clutter, 1963; Clutter et al., 1983; Vanclay, 1994).

Compatibility between growth and yield is the most important notion for producing growth and yield models, in that total growth should equal yield. An example of a compatible growth model is Matney and Sullivan's (1982) model for Loblolly pine stands. The example of this relationship is demonstrated using Schumacher's (1939) yield equation:

$$V=e^{\alpha+\beta/t} \tag{2.4}$$

First derivative of this equation provides a corresponding growth function:

$$\frac{dv}{dt} = -\frac{\beta}{T^2}\exp(\alpha + \frac{\beta}{T}) = -\frac{\beta}{T^2}V \tag{2.5}$$

2.3.1.3 Simultaneous Growth and Yield Model

Sullivan and Clutter (1972) further extended the notion of compatibility and produced a simultaneous growth and yield model. This approach expressed the future yields of a stand as a function of initial state variables and age such as basal area, site index and age, and has subsequently been referred to as the algebraic difference method (Borders et al., 1984). That is

$$Y_2 = f(Y_1, T_1, T_2)$$

where,

Y_2 = value of a stand variable at age T_2,

Y_1= value of same variable at age T_1,

T_1 = tree or stand age at initial time,

T_2 = tree or stand age at time 2.

When future age (T_2) equals initial age (T_1), the equation becomes a yield model, therefore, it is a simultaneously a yield equation for the initial condition and projection model for the future. Sullivian and Clutter's (1972) equation, a simultaneous equation, can be derived as follows (using IUFRO conventional symbols).

$$\ln(V) = \alpha_0 + \alpha_1 S + \frac{\alpha_2}{T} + \alpha_3 \ln(G) \tag{2.6}$$

$$\ln(G) = \beta_0 + \beta_1 S + \frac{\beta_2}{T} \tag{2.7}$$

and from 2.7

$$\beta_2 = T(\ln(G) - (\beta_0 + \beta_1 S)) \tag{2.8}$$

Differentiating 2.7 with respect to T, and substituting β2 as in 2.8 gives 2.9

$$\frac{d\ln(G)}{dT} = -\frac{(\ln(G) - (\beta_0 + \beta_1 S))}{T} \tag{2.9}$$

Integrating this equation and rearranging gives the form shown in 2.10

$$\ln(G_2) = (\frac{T_1}{T_2})\ln(G1) + (\beta_0 + \beta_1 S)(1 - \frac{T_1}{T_2})) \quad (2.10)$$

This equation represents a sigmoid and substituting In (G_2) in equation 2.6 produces final volume projection equation 2.11.

$$\ln(V_2) = \alpha_0 + \alpha_1 S + \frac{\alpha_2}{T_2} + \alpha_3 (\frac{T_1}{T_2})\ln(G_1) + (\alpha_4 + \alpha_5 S)(1 - (\frac{T_1}{T_2})) \quad (2.11)$$

where,

$\alpha 4 = \alpha 3\beta 0$

$\alpha 5 = \alpha 3\beta 1$

Sigmoid curves have been used to represent the yield of many biological growth processes (Causton, 1983). There are several sigmoid functional forms of growth and yield functions such as Schumacher, Chapman –Richards, Gompertz, Hossfeld, Weibull, monomolecular and others. Clutter et al. (1983) noted several desirable aspects of these functions:

1) representations of growth and yield should be compatible;

2) the functions should be consistent, i.e. as T_2 approaches T_1, Y_2 should approaches Y_1;

3) they should be path-invariant, i.e. predicting Y_3 from Y_1 should yield the same result as predicting Y_2 from Y_1 followed by Y_3 from Y_2;

4) as T_2 approaches ∞, Y_2 should approach an upper asymptote.

2.3.2 Diameter Distribution Models

Diameter distribution models provide estimates of yield per unit area by diameter classes. These models are useful for describing the properties and structure of a stand in terms of size classes, which is often required for effective management planning. Diameter distribution models occupy an intermediate position between the whole stand and individual tree models in terms of description detail, computational cost, and information requirements (Garcia, 1988).

Probability density functions have been key procedures in generating diameter distributions, and several probability density functions have been used successfully, such as the Weibull distribution (Bailey and Dell, 1973), the Gamma distribution (Nelson, 1964), the Beta distribution (Clutter and Bennett, 1965) and Johnson's S_B distribution (Hafley and Schreuder, 1977). The Weibull distribution has been the most extensively used function in recent diameter distribution modelling, because this function provides the following desirable properties:

1) it is relatively simple to manipulate mathematically;

2) it has flexibility;

3) it has a closed form (Bailey and Dell, 1973).

The Weibull is a three-parameter distribution defined by the probability density function:

$$f(X) = \frac{c}{b}\left(\frac{X-a}{b}\right)^{c-1} \exp\left[-\left(\frac{X-a}{b}\right)^{c}\right] \qquad (a \le X < \infty)$$

$$= 0, \text{ otherwise} \qquad (2.12)$$

where,

X=dbh,

a = location parameter,

b = scale parameter,

c = shape parameter.

The cumulative distribution function for the Weibull distribution is

$$F(X) = 1 - \exp\left[-\left(\frac{X-a}{b}\right)^{c}\right] \qquad (a \leq X < \infty) \qquad (2.13)$$

$$= 0\,, \text{ otherwise.}$$

For application to the distribution of diameter at breast height (dbh), the scale (*b*) parameter and the shape (*c*) parameter should always be positive, and location (*a*) should be positive or zero. Parameters of the density function were estimated directly as functions of stand variables (for example, age, stems/ha and site index), therefore, values of the stand variables could be estimated by integration of predicted probability density functions. This method, however, has been found to produce inconsistent estimates of whole stand as compared to simultaneous growth models (Whyte and Woollons, 1992).

The parameter recovery method involves the recovery of Weibull distribution parameters from models of basal area/ha, dbh variance and maximum dbh, and resulted in compatible basal area and distribution models (Hyink and Moser, 1983; Bowling et al., 1989).

Kuru (1989) and Kuru et al. (1992) found out that using a reverse Weibull distribution was a more effective way of representing diameters than using the minimum diameter, because it showed a stronger relationship with time. Liu Xu et al. (1992) used an

extreme value distribution to adjust maximum estimates for Douglas-fir plot size, and the system worked well for Caribbean pine in Fiji (Villanueva and Whyte, 1992).

Theoretical problems with diameter distributions were revealed. For example spatial correlation of the diameter in a plot by Garcia (1991) could cause a distribution to vary depending on the land size in consideration. Whyte and Woollons (1992) suggested that conducting more comprehensive routine inventory that is properly coordinated with growth modelling and yield forecasting, should solve the problem.

On the other hand, the relative basal area method allows a distribution model to be recovered from a basal area/ha model. Relative size of the i^{th} tree is defined as $g_i/\bar{g}$ where g_i is the basal area of the i^{th} tree and $\bar{g}$ is the mean basal area per tree (Clutter and Allison, 1974; Woollons and Hayward, 1985; Pienaar and Harrison, 1988; Pienaar, 1989). This method considers the effects of allocating growth to the various diameter classes which are defined on the basis of relative size. An initial hypothesis, was that the relative size of survivor trees remains constant over time. However it also implies that the i^{th} survivor contributes a constant proportion of the total basal area over time, so that the future size of the i^{th} survivor, g_{2i}, can be obtained from its present proportion of total basal area of survivors, $g_{1i}/\bar{g}$. In reality, the relative size decreased over time with the contribution of smaller than averaged-sized survivors, whereas the relative size of the largest trees increased over time (Pienaar and Harrison, 1988; Pienaar, 1989). So, a different model of R is required:

$$R_{2i} = R_{1i}^{\left(\frac{\tau_2}{\tau_1}\right)^{\alpha}} \qquad (2.14)$$

where,

$R_{2i} = g_{2i}/\bar{g}_2$;

$R_{1i} = g_{1i}/\bar{g}_1$;

$\overline{g}_1$ = average per tree basal area at age T_1;

$\overline{g}_2$ = average per tree basal area at age T_2;

T_1= stand age at the beginning of the growth period;

T_2= stand age at the end of the growth period;

α= parameter to be estimated.

Once an estimated value of α is available, a projected stand table consistent with the observed or projected total basal area G_2 can be obtained as follows:

$$n_i g_{2i} = G_2 (g_{i1} / \overline{g}_1)^{\beta} n_{i1} / \sum_{i=1}^{k} (g_{i1} / \overline{g}_1)^{\beta} n_i \qquad (2.15)$$

where $\beta=(T_2/T_1)\alpha$, n_i is the number of survivors in the i^{th} initial dbh class (I=1,2, . . ., k), and g_{2i} the projected total basal area of these n_i survivors.

When total survival and projection equations are available, the proposed stand table projection procedure requires that the predicted total mortality be identified in the initial stand table. For this purpose it is assumed, that the probability that a given tree will die during the projection interval is inversely proportional to the square of its relative size. Given that a tree dies, it is then possible to compute a probability of class mortality by partitioning the predicted total stand mortality and allocating the components among the various diameter classes.

Despite these perplexing considerations, diameter distribution approaches to growth and yield modelling have remained useful methods in modelling growth.

2.3.3 Individual Tree Growth Models

In alternative approach, individual trees are identified and individual tree characteristics are used as the basis for modelling. A list of attributes, for example: species, dbh and stem defects are recorded for each individual tree. Individual tree models have been classified as distance-independent and distance-dependent tree level models (Munro, 1974). Individual tree models could be used when stand structure is so irregular that density cannot be reasonably described by a single measure for the entire stand (Bruce and Wensel, 1987).

Distance-dependent models use distance to adjacent trees as independent variables, while in contrast distance-independent models do not use distance to adjacent trees as independent variables for predicting the growth of individual trees. The growth of an individual tree is a function of variables such as present size, site index and stand density in distance-independent models. The mortality tree level models may be predicted from past growth rates or generated randomly. Lemmon and Schumacher (1962) modelled volume and diameter growth of ponderosa pine as a function of size, age, density and site index. PROGNOSIS (Stage, 1973) offers great potential, particularly through its designed ability to function as a feedback mechanism for localisation of general models.

Distance dependent models are based on the assumption that individual tree growth can be predicted more precisely if sizes and locations of neighboring plants are known (Clutter et al., 1983). Thus, they provide very detailed information about the structure of a stand. Variables for each tree are calculated as a function of its size, site quality and some function of distance to and size of neighbouring trees, which serves as an expression of competition. This type of model has been produced by many researchers (e.g. Arney, 1974; Hegyi, 1974; Lin, 1974; Leary, 1979; Tennent, 1982; Nystrom and Kexi, 1997).

Bruce and Wensel (1987) pointed out that using a more complicated model than necessary, often has two costs; 1) greater computational expense and 2) a loss of precision of estimates.

2.3.4 Summary

Three growth and yield modelling approaches, which were classified by Munro (1974) have been discussed. These included whole stand, diameter distribution and individual tree models. None of these approaches can be regarded as the ideal one. The right model to choose depends on several important factors (Bruce and Wensel, 1987):

1) the application of resources used. It is the end use that finally determines the best approach for forest stands;

2) models that are unnecessarily complicated may incur several costs such as greater computational costs, loss in precision of estimates, and difficulty in understanding and assessing the utility of the model.

Buckman and Shitleg (1983) offered a more detailed checklist to help potential users assess the suitability of a growth model for stated application. Key items include ease of use, accuracy of predictions, and biological realism. The appropriate approach depends on the data used, facilities and expertise available. Thus, the model should be objective, unbiased, documented and available.

2.4 Site Quality

Estimates of site productivity are an important requirement for predicting forest growth and yield. Some sites produce luxuriant forest, while others are capable only of supporting poor forest. These differences may be because of soil (e.g. fertility, density),

climate (e.g. temperature and rainfall patterns), topography (e.g. elevation, aspect) and other factors, and may be reflected in the species composition and growth patterns. (Vanclay, 1994). Methods of classifying site quality entail the use of direct and indirect methods, which are discussed below.

2.4.1 Indirect Methods for Estimation of Site Quality

Indirect evaluation of site quality includes the use of indicator species, and the prediction of fertility from environmental variables. Such usage depends on knowledge of the relationships between certain growth patterns of the species of interest and growth patterns of the species available for measurement (Clutter et al., 1983).

The composition of understorey species has been used to provide information on site quality for tree growth. However, the application of this approach has some drawbacks:

1) the method permits site evaluation on relative or qualitative terms;

2) the understorey characters are generally influenced by disturbing factors such as fire, wild life, grazing and site preparation treatment;

3) understorey vegetation reflects only the fertility of the topmost horizons of the soil profile, and very little about deeper horizons which greatly influence site quality for tree growth.

Examples of site evaluation using this method have been reported by Ure (1950), Spurr and Barnes (1980), and Philip (1994). Ure (1950) related site productivity in Kaingoroa forest to the presence or absence of certain understorey plant species. Inter-species relationships also have been used in evaluation of site index. Coile (1948) used this method to calculate site index of loblolly pine and short leaf pine, where both species were present. Regression was calculated using short leaf pine site index as the

independent variable and loblolly pine site index as the dependent variable. However because of limitations and complexities of applying the indicator species method and no available data, it will not be used in this study. Instead site quality estimation will be based on stand height data, which is by far the most commonly used technique for evaluating site productivity.

Other indirect methods for estimation of site quality use topographic, climatic and edaphic factors as independent variables. For example site index of radiata pine in the North Island of New Zealand has been related to mean annual rainfall and temperature at the closest meteorological stations, plus soil nutrients, topsoil depth, soil penetrability and soil pH (Hunter and Gibson, 1984).

2.4.2 Direct Methods for Site Evaluation

The direct estimation of site quality can be classified into two major categories: 1) historical yield records, and 2) site index. Historical yield records could be used to directly evaluate site index in terms of a measure of prediction in physical quantities. However, historical data are usually limited and even when one or more rotations have been observed factors such as species composition, changes in site preparation techniques, rotation age, and genetics may reduce the utility of site quality evaluation based on the history of previous rotations.

In theory measuring site quality by the volume of desired material produced, thus expressing the integrated net effect of all site factors in terms of the product itself, would be the most logical method in determining site productivity. In practice however, total stand volume fluctuates due to, changes in stand density and the implementation of silvicultural practices. Stand density is often determined by the silvicultural regime which is influenced by site quality. At any given age, density in an unthinned stand might be expressed as stems/ha, but measures which relate numbers of stems to average tree size are generally more independent of age and site quality.

The silvicultural practice of pruning causes significant impacts on tree growth. Pruning live branches generally reduces the amount of foliage and growth potential of the tree and stand. Thus total stem volume is usually expected to drop following pruning. Although pruning prescriptions have largely focused on the amount of crown that can be removed while minimising the loss in volume growth, the main objective is to maximise wood quality. Even though pruning slows tree diameter growth, the economic benefits of improved wood quality usually counteracts increases in wood quantity. Site quality is therefore rarely measured directly through stand volume, except when the factors which affect volume prediction have been strictly controlled (Clutter et al., 1983). Nevertheless this method is not practical because in order to control those factors, high costs are involved.

The most frequently used representation of site productivity is site index, which is the expected top height at a specified index age (Avery and Burkhart, 1994). Theoretically height growth is sensitive to differences in site quality, slightly affected by varying density levels and species composition, relatively stable under varying thinning intensities, and strongly correlated with volume. For species in which height growth is significantly influenced by modern stand density variation, estimation of site quality from stand height data will provide poor results unless the effect of stand density is taken into account. Fortunately, for many important plantation species, height growth is rarely affected by variation in stand density (Assmann, 1970; Lanner, 1985). Consequently site quality estimation procedures based on stand height information, are the most commonly used techniques for evaluating site productivity and this approach was therefore used in this study.

Site quality evaluation using height growth, involves development of site index curves which are a family of growth patterns. Many previous studies have used this approach such as Brickell (1968); Beck (1971); Carmean (1972); Bailey and Clutter (1974); Burkhart and Tennent (1977); Clutter and Jones (1980); Borders et al. (1984); Bailey et al. (1989); Temu (1992); Ngugi (1996).

2.4.3 Development of Site Index Curves

Site index estimation can be developed from three sources:

1) measurement of stand height and age in temporary plots;

2) measurement of stand height and age in permanent sample plots;

3) individual tree stem analysis.

There are three usual methods for developing site index curves or equations (Clutter et al., 1983). These are:

1) guide curve method;

2) difference equation method;

3) parameter prediction method.

Site index curves can be classified into three types: 1) anamorphic curves; 2) polymorphic-disjoint curves; and 3) polymorphic-nondisjoint curves according to the nature of the height/age curve families they generate (Clutter et al., 1983; Borders et al., 1984). Anamorphic curves have the shape parameter (β) eliminated, constraining the resultant curves to have similar shape, but with different asymptote parameters (α). Thus, for any two curves in an anamorphic family, the height of one curve at any age is a constant proportion of the height of the other curve at the same age. Polymorphic equations have the shape parameter (β) which allows each curve the freedom to change in shape, while asymptote parameters (α) may, or may not vary over the curves in the family (Rennolls, 1995). Hence, there is no constant proportionality like in anamorphic curves, but the curves do not cross each other within the age of interest in a

polymorphic-disjoint family curve. While in polymorphic-nondisjoint family curves, one curve can cross another within the range of interest.

The major weakness of anamorphic curves is the assumption of a common shape for all site classes. For some species, the height curve shape varies with site quality. Higher quality lands generally exhibit more pronounced sigmoid shapes and lower quality lands produce flatter height growth patterns. Polymorphic families of site index curves produce differing shapes for different site index classes (Avery and Burkhart, 1994). Thus, polymorphic curves generally reflect height growth trends across a wide range of site qualities more accurately than anamorphic curves.

2.4.3.1 Guide Curve Method

The guide curve methods generate site curves by graphical techniques and regression analysis (Alder, 1980). Data on height and ages were collected from various stands on different sites and of varying ages. These paired height-age values were then plotted and a guide curve was drawn to depict the general trend in data. All other site index curves were then proportional to the guide curve. Anamorphic site index curves have generally been constructed by this method.

2.4.3.2 Difference Equation Method

The difference equation method for site index usually requires permanent sample plots or stem analysis data. The procedure of formulating a difference equation is flexible and can be used with height and age equations to produce anamorphic and polymorphic families curves (Clutter et al., 1983). A procedure of developing a difference equation is shown using the Schumacher log reciprocal function below.

A modified Schumacher height equation is given as:

$$\ln(H) = \alpha + \beta/T^{\gamma} \tag{2.16}$$

where,

H = height in meters

T = age in years

α, β and γ = coefficients to be estimated.

The heights are given by equations 2.17 and 2.18 at time T_1 and T_2:

$$\ln(H_1) = \alpha + \beta/T_1^{\gamma} \tag{2.17}$$

$$\ln(H_2) = \alpha + \beta/T_2^{\gamma} \tag{2.18}$$

Making β the subject in equations 2.17 and 2.18, and equating the two resultant equations give equation 2.19:

$$T_2^{\gamma}(\ln(H_2)-\alpha) = T_1^{\gamma}(\ln(H_1)-\alpha) \tag{2.19}$$

Solving this equation for $\ln(H_2)$, gives equation 2.20:

$$\ln(H_2) = \ln(H_1)^{(\frac{T_1}{T_2})^{\gamma}} + \alpha(1-(\frac{T_1}{T_2})^{\gamma} \tag{2.20}$$

taking the exponent of equation 2.20 gives equation 2.21:

$$H_2 = H_1^{(\frac{T_1}{T_2})^{\gamma}} \exp(\alpha(1-\frac{T_1}{T_2})^{\gamma}) \tag{2.21}$$

When T_2 in equation 2.21 is set equal to an index age (e.g. 40 years for Douglas-fir in New Zealand), then H_2 will be an explicit definition of site index.

2.4.3.3 Parameter Prediction Method

The parameter prediction method of fitting site index curves has been used by many researchers for fitting polymorphic-disjoint site index equations. This method requires data from permanent sample plots or stem analysis data and involves the following steps (Clutter et al., 1983):

1) fitting a linear or non-linear height function to the data from each tree or plot;

2) assigning a site index value to each tree or plot using the fitted value; and

3) relating the parameters of fitted curves to site index through linear or non-linear regression procedures.

2.4.3.4 Limitations of Site Index

There are several possible limitations of the use of site index as a measure of forest productivity (Monserud, 1985, 1987; Mason, 1992; Avery and Burkhart, 1994):

1) site index is not a constant, it may change periodically because of environmental and climatic variations;

2) height is usually measured with less precision than other stand state variables such as basal area/ha;

3) the site index value for one species usually cannot be translated into a usable index for different species on the same site (Doolittle, 1958).

Despite these limitations, site index is a useful indicator of site quality because it provides a simple numerical value that is easily measured and understood by practicing foresters.

2.5 Localising Growth and Yield Models

General growth projection systems have often been developed to cover large geographic forest areas or regions. Examples of these include PROGNOSIS (Stage, 1973), STEMS (Belcher et al., 1982), SIDFIR (Law, 1990). However because of their broad development, some potential exists that these models will not provide adequate sensitivity of estimation for sub-regions. This is because unexplained factors within sub-regions can be averaged for the whole area, but not within sub-regions, resulting in biased estimates.

When intensive forest management demands growth predictions that are sensitive at regional or sub-regional levels, then general models lose their credibility and growth models will be restricted at sub-region level (Whyte et al., 1992).

Various methods have been used for localising regional models to sub-regions. The method of stratification involves modelling each different stratum individually. To justify this hypothesis, all differences between parameters of each stratum must be conduced. Burkhart and Tennent (1977) used this method to fit site index equations for radiata pine grown in New Zealand. Other methods have been used to localise growth models. Smith (1983) used the double sampling technique of Cochran (1977), to calculate an annual adjustment factor of diameter increment of the STEMS model (Shifley and Fairweather, 1983). The Bayesian method of estimating model coefficients has also been used by Berkey (1982) and Green et al. (1992). Green et al. (1992) reported a reduction of more than 50% in residual mean squares (RMS) by simultaneously estimating Honduran pine yield equation coefficients, for sub-populations with 21 different soil site groups. Gertner (1984) used a sequential Bayesian

method which adjusted the parameter estimates through time to localise a diameter increment model taken from STEMS.

Dummy variables have also been used to localise growth and yield models. These variables are used to convert qualitative information such as geographical region and season of the year into quantitative information by means of a coding scheme (0 or 1). This method involves formulation of an analysis of covariance among regions or data sets, by representing each as a dummy variable within a single equation. Regressing Y on an intercept dummy variable (X_1) in addition to one or more measurement variables (X_2) yields equations with form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \qquad (2.22)$$

- the Y-intercept equals β_0 when $X_1 = 0$
- the Y-intercept equals $\beta_0 + \beta_1$ when $X_1 = 1$
- β_1 equals the difference in Y-intercept between X_1 categories 0 and 1

A test of H_0: $\beta_1 = 0$ therefore determines whether the two intercepts differ significantly. Intercept dummy variables test for differences in intercepts. To test for a difference in slopes, an interaction term called a slope dummy variable is forms by multiplying the dummy times measurement variable. Regressing Y on both a measurement variable X_2 and a slope dummy variable X_1X_2 yields equation with form

$$Yi = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} \qquad (2.23)$$

- the slope relating X_2 to Y equals β_2 when $X_1 = 0$
- the slope relating X_2 to Y equals $\beta_2 + \beta_3$ when $X_1 = 1$
- β_3, the coefficient on X_1X_2, equals the difference in slopes between X_1 categories 0 and 1

A test of H_0: $\beta_3 = 0$ determines whether the two slopes differ significantly. Although the above example applies to linear regression, the same principles are applicable to non-linear models. The use of dummy variables thus provides a potential to test whether or not different models for different sub-populations are justified.

Gujarat (1970) has demonstrated the general approach of this method, and Ferguson (1979) used 4 dummy variables to localise a basal area increment model for 5 forests, mainly to represent different rainfall patterns. Whyte et al., (1992) and Temu (1992) also used dummy variables to model growth of Douglas-fir in the South Island of New Zealand, and Mason (1992) for radiata pine initial growth models in the North Island of New Zealand. The use of dummy variables thus provides potential capabilities for testing the justification of having different models for different sub-populations, where a number of site variables such as altitude, latitude, distance from the sea, annual rainfall and soil type can be considered as possible variables to explain growth variation across regions.

2.6 Data Collection and Preparation for Growth and Yield Modelling

Inventory is used for many purposes, but different procedures are required to satisfy various needs of different users in an efficient way. Permanent sample plots provide the basis for growth modelling, yield prediction and sustained yield management, and the reliability of these data is crucial to these and many other aspects of forest management. Vanclay (1994) pointed out some key points which are necessary to obtain reliable data. They were:

1. ensure consistent standards;
2. sample a wide range of stand and site conditions;
3. provide both passive monitoring and experimental plots;

4. check that measurement records are unambiguous and secure.

Satisfactory growth models may depend on the availability of high quality data from a wide range of stand conditions and treatments. One of the main principles in collecting the data for growth and yield modelling is to sample the full range of site and stand conditions. The need to provide reliable data for growth and yield modelling demands three qualities of permanent sample plots that are not necessary in continuous forestry inventory and other permanent plot systems (Vanclay, 1994). These are: 1) individual trees must be unambiguously identified; 2) the plots should be homogeneous; and 3) the plots should sample extremes of site and stand conditions.

Trees should be identified using permanent marks and unique numbers. This not only offers more options for modelling, but also is the only sure way of detecting measurement errors. Growth modelling requires homogeneous plots which means minimising the plot variance. The ability of the permanent plots to quantify the present resources is irrelevant. Thus the same plot series cannot be efficiently used for both resource inventory and growth and yield model development. If the growth model is to be used to investigate silvicultural and management alternatives, the database must include experimental data with paired treatment and control plots both with adequate isolation. In contrast to continuous forestry inventory plots, it is not necessary for the permanent plots to be representative or numerically proportional to forest areas, but it is essential that they sample the full range of stand conditions.

Data resulting from re-measured permanent sample plots or trees, is referred to as a real growth series. Real growth series data are available for many tree species in plantations. Such real growth series data have extensively been used for investigating tree and stand dynamics as well as for modelling growth. An often used growth modelling formulation was developed by Clutter (1963), and has been referred to as the algebraic differential equation (ADE), (Borders et al., 1984; Ramirez et al., 1987; Villanueva, 1992).

In using projection equations of this functional form, it was appropriate that real growth series data available from a system of PSP for Douglas-fir plantations in the South Island, be used to obtain sample estimates of the parameters of equations that best described the growth and yield of the selected stand variables.

Real growth series data can be used in fitting routines to obtain sample estimates of population parameters for ADE's. When tree or plot measurements have been recorded at only two ages, T_1 and T_2, obviously the only growth interval is from T_1 to T_2. However if tree or plot measurements have been recorded 3 or more times, as is usually the case of most permanent sample plots, then two more growth intervals can be derived. For example, if a real growth series for a set of units (tree or plot) is measured at ages $T_1, T_2, T_3, \ldots, T_n$, there are $\binom{2}{n}$ possible growth intervals that can be derived, as well as the longest interval for each unit (i.e. T_1 to T_n). The most common procedure for estimating parameters in algebraic differential equations is to use nonoverlapping growth intervals (Clutter, 1963; Sullivan and Clutter, 1972; Borders et al., 1984).

However the use of all possible combinations of data intervals has been observed as a powerful approach in previous studies (Villanueva, 1992 and Chikono, 1994). What has not been obvious, is which growth intervals should be used to obtain the parameter estimates. Thus, in this study the following three data sets were derived from all possible interval data sets. Short interval, long interval and mixed interval formulations were compared to investigate a precision of coefficient estimations and thus model effectiveness can be increased.

2.7 Douglas-fir in General

Douglas-fir grows naturally in most of the temperate portions of western North America, where it is one of the most valuable and commercially important timber species. The natural range of Douglas-fir occurs in a board coastal band in America which lies to the west of the coast range in British Columbia, the Cascade Range in Washington and

Oregon, and the Sierra Nevada on the northern Californian coast. It's range extends from latitude 55° north in British Columbia to latitude 35° north on the Californian coast (McArdle et al., 1961; Oliver et al., 1986; McComb et al., 1993; Miller and Knowles, 1994). Douglas-fir grows in regions which have generally mild climates that are humid except for a dry period in summer. Throughout its latitudinal range it occurs close to sea level, but it also rises into the mountains reaching altitudes of about 800 metres above sea level in British Columbia and 1800 metres above sea level in California, where there is a higher rainfall ranging from 920 to 3200 mm (McArdle et al., 1961; Miller and Knowles, 1994).

The species reaches its best development in North America on deep, well-drained soils where annual rainfall is approximately 1750 mm, and the climate is not subject to extremes of temperature. Here individual trees can grow to over 4 metres in diameter and 100 metres in height (Barrett, 1980; Oliver et al., 1986).

Douglas-fir is regarded as moderately shade tolerant. Young Douglas-fir seedlings require only about one-third the natural sunlight intensity to achieve maximum photosynthesis under optimum temperature and moisture conditions. On severe sites approximately 50 percent shade produces the best survival and reasonable growth of seedlings (Burns, 1983). Shade tolerance varies with age and site conditions. Young trees are more shade tolerant than older trees and are able to tolerate side shading, while growth of trees over 25 years of age is reduced by either side or overhead shade (McArdle et al., 1961).

2.7.1 Douglas-fir in New Zealand

2.7.1.1 Distribution of Species

Douglas-fir makes up almost 4 % of New Zealand's plantation forest, and is the second most common species after radiata pine. By 1996 approximately 70,000 ha were

established in Douglas-fir with just over half of this resource located in the central North Island and the remainder in the South Island, mainly in Nelson, Canterbury and Southland. Douglas-fir has been planted at a relatively steady rate averaging 1500 ha per year for the last 30 years. Since 1974 the planted area has increased substantially, particularly in the central and southern North Island regions and in Nelson, Canterbury and Southland (Neilson and BMS Ltd., 1994; Knowles et al., 1995).

2.7.1.2 Site Requirement

Douglas-fir requires an annual rainfall greater than 800 mm per year and grows well from the Central North Island to Southland, particularly in areas which receive moderately high rainfall of 1000 to 1500 mm annually. However, it has not thrived in the north of the North Island. This has been attributed to a combination of warmer conditions, unsuitable provenances, heavy clay soils and exposure to salt-laden winds (Miller and Knowles, 1994; Knowles et al., 1995). Douglas-fir performs best on fairly deep, well-drained soils of at least moderate fertility. Suitable sites also need to be relatively free of competing vegetation, as initial growth after establishment can be slow (Knowles et al., 1995).

Spurr (1961) suggested that altitude limits for good growth in Douglas-fir, were 900 m above sea level in the North Island and 750 m in the South Island. However, Miller and Knowles (1994) pointed out that these estimates were conservative because Douglas-fir is growing at commercially productive rates well above 800 m in the Craigieburn Range in Canterbury and up to 920 m at Karioi in the central North Island. Knowles et al. (1995) reported that upper altitudinal limits are approximately 975 m above sea level in the North Island and 850 m above sea level in the South Island.

For this study altitude ranges within the data set varied from 0 to 790 m above sea level. These altitude ranges lay within the limits that were suggested for good growth of Douglas-fir by Knowles et al. (1995).

2.7.1.3 Silviculture and Management

Douglas-fir has traditionally been grown to provide specific timber markets for framing of engineering purposes. These are end uses in which the critically important characteristics are stiffness and strength, factors that are strongly related to silviculturaly controllable features such as branch size. More recently in New Zealand, there has been a strong demand for export logs, including pruned logs of Douglas-fir.

Silvicultural practices prior to 1970 involved initial stocking levels that were relatively high 2500-3000 stems/ha. This resulted in high quality timber with small branch size. But this involved high establishment costs, and required long rotations of 60-80 years. Also thinning was delayed until stands were 30 and 40 years of age (Miller and Knowles, 1994). Many stands have been selectively pruned in the past, but little is known about the effects of pruning on growth and log quality. Due to Douglas-fir producing a very large number of branches, even though many are small, pruning costs are twice as much per tree as for radiata pine. In the past (1960s and 1970s), many stands that were around 30 years of age received a single production thinning, yielding 90-150m^3/ha as small sawlogs. Current practice involves a waste thinning before age 20 to a stocking of 300-600 stems/ha with the intention of no further thinning, or to 500-800 stems/ha if there is to be a later production thinning. This production thinning is usually scheduled at age 25-30 years with the stand being reduced to 250-500 stems/ha (Knowles et al., 1995).

Because of the radically changing economic climate, past silvicultural practices which have provided the structural timber now on stream would today result in prohibitive growing costs. Current silvicurtual regimes aim to provide the desired quality of timber at a reasonable cost. These regimes require a much shorter rotation length of about 45 years, and lower initial stocking rates of approximately 1500 to 1600 stems/ha. Early thinning is essential to maintain diameter growth and promote a healthy green crown and should result in final stockings of 250 to 500 stems/ha (Miller and Knowles, 1994).

2.7.1.4 Uses of Douglas-fir

Douglas-fir is highly regarded as a commercial timber species throughout the world, in part due to its fast growth relative to many other species. In the past the chief sources of timber has been the natural forests of the American Pacific Northwest, usually referred to as "old growth". Today this resource is decreasing, and is being set-aside for conservation and environment preservation purposes. Its place is being taken by younger planted material, the so-called "second growth" Douglas-fir, which New Zealand Douglas-fir can be subsequently termed.

In the past old growth stands of Douglas-fir in America have been a major source of structural timber, veneer, plywood and pulp and paper (McArdle, 1961). Because of its good strength and stability, New Zealand grown Douglas-fir is primarily suited for use as a structural and framing timber. It has greater fibre length than radiata pine, and a major advantage is that wood density and strength do not decrease near the pith, allowing framing timber to be sawn from much smaller logs such as thinning. As a structural timber, Douglas-fir provides better recoveries than radiata pine and is an excellent timber for use as beams and columns. According to Cown (1992) as high as 60% of the recoverable timber can be used for structural and engineering purposes when it is stress-graded, and does not need drying and preserving for interior use.

Douglas-fir has also been an important pulpwood species in the Pacific Northwest region of the USA. It is especially valued for the improvement it brings in tear strength of paper products when mixed with pulp made from other species. Although its use as pulpwood in New Zealand is limited, kraft pulps are easily prepared from New Zealand grown Douglas-fir, at slightly lower yields than radiata pine. It has a high tear index relative to radiata pine. The chief constraint on its use is the high bleach requirement, arising from its dark heartwood colour (James, 1975;Cown, 1992).

In New Zealand, very little interest has been expressed in the use of Douglas-fir in Veneer and plywood manufacture, despite the fact that it is the mainstay of the Pacific Northwest industry. Research trials have shown that local logs peel poorly and, compared with radiata pine, veneer characteristics are highly variable due to the contrasts in earlywood and latewood (Miller and Knowles, 1994). Douglas-fir has a well established role as an amenity and shelter species. Without browsing, open grown specimens will often branch almost from ground level, allowing Douglas-fir hedges and shelterbelts to provide a satisfactory screen within a few years of establishment (Miller and Knowles, 1994).

2.7.1.5 Marketing of Douglas-fir

Douglas-fir has an excellent reputation as a general purpose structural timber on both domestic and export markets. The long established reputation of "Oregon pine", the commercial trade name for timber originating from the native stands in North America, provides a key marketing advantage to all Douglas-fir timbers. Huge shortfalls in the supply of Douglas-fir to traditional markets are already occurring because of reduced availability in the American Pacific Northwest due to increasing conservation pressure to preserve habitats for wildlife such as the spotted owl. Less than 18% of the original old-crop Douglas-fir resource remains to be harvested (Miller and Knowles, 1994). In the future, timber from "second-growth" forests including New Zealand Douglas-fir will satisfy the market.

Douglas-fir exports from New Zealand include both sawn timber and logs. The three largest markets for New Zealand exports of sawn timber from 1991 to 1993 were Australia, New Caledonia and Japan respectively. Douglas-fir is also currently the major softwood log species imported into the Japanese market, comprising 23% of total log imports and 36% of the Softwood log imports (Miller and Knowles, 1994). Most of this timber still originates in North America. New Zealand recently adopted the same log

grading system as American producers and is making a significant contribution to supplies.

Douglas-fir has a well established reputation as a commercial softwood species suitable for a variety of New Zealand sites. Miller and Knowles (1994) suggested that the following developments need to be implemented to improve the productivity and economic gain of New Zealand Douglas-fir:

1) the development of a sound genetic improvement programme;

2) an improved understanding of the effects of site and silviculture on stand growth and timber quality;

3) the acceptance in the international marketplace of a reduced quality product compared with the "old-growth" Douglas-fir.

2.8 Growth and Yield Models for Douglas-fir in New Zealand

The first Douglas-fir growth and yield model in New Zealand was DFIR developed by Mountfort (1978), for the Kaingaroa forest in the central North Island. A second model was developed for the Nelson region by NZFRI to cater for production of Douglas-fir in the Nelson area. Before these models, management of Douglas-fir plantations relied on conventional yield tables prepared by Duff (1956), Spurr (1961, 1963) and Elliott (1969). The first complete Douglas-fir models in New Zealand were Douglas-fir Central North Island Growth Model Version 1 (DFCNIGM1) and Douglas-fir Central North Island Growth Model Version 2 (DFCNIGM2), of which both are whole stand models derived by Liu Xu (1990). Liu Xu also developed Douglas-fir Central North Island Growth Model Version 3 (DFCNIGM3) in 1990, which was a diameter distribution

model compatible with DFCNIGM2. The DFCNIGM1 and DFCNIGM2 models endeavor to model the effect of Swiss needle-cast fungus.

DFCNIGM1 is a whole stand model for Central North Island plantations that included Kaingaroa, Pureora, Waimihia and Whirinaki. This model consisted of three parts: a growth and yield projection for 1) healthy, 2) diseased and thinned, and 3) diseased and unthinned stands. Use of this model by the New Zealand Forestry Corporation, proved to provide satisfactory prediction for all stand statistics, except the mortality equation which appeared to provide overestimation. DFCNIGM2 revised the mortality function of DFCNIGM1, and also replaced the basal area projection equation originally derived from the Schumacher equation, with the Hossfeld equation (Liu Xu, 1990).

DFCNIGM3 is a diameter distribution model for the same plantation resources. The reverse Weibull distribution was used as the probability density function for characterising the distribution of diameters at breast height over bark (dbhob) using maximum dbhob as a location parameter. The extreme value distribution was used to adjust size class range. Whyte and Woollons (1992) pointed out that even though more theoretical work is needed on this extreme value distribution, the empirical evidence from this study indicates the practicalities of this approach performed well.

Law (1990) developed the South Island Douglas-fir growth model (SIDFIR), which is a whole stand growth model. Data used in the construction of the model originated from forests throughout the South Island, ranging from the Nelson area down to Southland. Law followed the method that was based on a state-space model, which was advocated by Garcia (1979,1984,1987, 1988, and 1994), comprising a set of stochastic differential equations. The rates of change of the stand parameters of mean top height, stand basal area, and stems/ha were predicted over time on the assumption that future development is determined by the values of the stand at any time. To be more specific, the state at a given time, t, specified by a list of n numbers (state variables), is explained by an n–dimensional state vector x(t). The inputs and outputs are also finite-dimensional vectors

u(t) and y(t), respectively. Then the behavior of the system is described by a transition function:

$$X(t) = F[x(t_0), u, t-t_0) \tag{2.24}$$

And an output function:

$$Y(t) = g[x(t)] \tag{2.25}$$

Equation 2.24 gives the state at any time t, as a function of the state at some other time t_0, of the inputs denoted by u, and of the elapsed time between t_0 and t. The output function 2.25 estimates outputs as a function of the current state. Transition functions are generated by integration of differential equations. The model can then be stated as:

$$\frac{dx}{dt} = f(x,u) \tag{2.26}$$

$$y = g(x)$$

The coefficients of the system are simultaneously estimated using maximum likelihood techniques.

Another model for Douglas-fir grown in the South Island is DfirStand which was developed by Temu (1992). DfirStand was a whole stand simultaneous growth and yield model applicable to four regions of the South Island. Equations employed by DfirStand showed an acceptable level of precision, but unlike other regional models they did not provide overall average estimates. In contrast to the approach adopted previously in SIDFIR (Law, 1990) which predicts overall average values for the whole of the South Island, DfirStand attempted to provide flexibility for aggregation and disaggregation by employing dummy and local site variables, along with other predictor variables to improve the sensitivity of estimation. However, DfirStand provided local variation for

only the basal area/ ha equation. Temu also developed tree volume and taper models collectively called DfirTree. DfirTree is a compatible tree volume and taper prediction system, developed to provide estimates of tree volume and stem diameters from inputs of *d* and *h* for Douglas-fir trees grown in the South Island. The volume-taper prediction system was based on the principle of splines (segmented polynomials) and provided two approaches with which to determine volumes of any part of the stem: 1) volume based and 2) taper based predictions. Preliminary work in this study found that there was no significant difference in terms of accuracy and precision between these two approaches and that precision was far superior in relative terms to that in the growth projection equations. Consequently, this study put emphasis on growth projection equation which is of greater priority rather than modelling stem volume and taper equations.

Examinations of the validity of DFCNIGM2, SIDFIR and DfirStand growth models were tried by Tennent (1995), who reported that all three models predicted growth within generally acceptable limits, and of course the models performed best in the regions from which their data were originally derived from.

Chapter 3

Data and Data analysis

3.1 Available Data

The data used in this study were obtained from re-measurements of permanent sample plots maintained by Forest Research Institute of New Zealand, Rotorua, after having sought and gained permission from private companies who own or lease regional forests.

The data set included measurements from four forests growing regions: the Canterbury set comprised 57 plots from 4 forests; Nelson 136 plots from 3 forests; Southland 128 plots from 8 forests; and Westland 45 plots from 5 forests. Basic and derived variables came from 366 permanent sample plots for this study. Plot sizes ranged from 0.01 hectare to 0.2 hectare with a mean size of 0.07 hectare. Trees per plot were from 8 to 502, with a mean of 72 trees per plot. Table 3.1 shows the distribution of plots by regions.

In the Canterbury region, mean age of trees was 29 years, minimum and maximum ages were 9 and 57 years respectively. Distributions of site index at index age 40 were from 16.4 m to 31.9 m and the mean was 26.5 m. Altitude was distributed from 150 m above sea level to 790 m, with a mean of 330 m. Stems per hectare ranged from 74 stems to 2970 stems, while the mean was 1033 stems per hectare. In the Nelson region, ages of

trees were from 7 to 60 years, with a mean of 28 years. Altitude and site index were distributed from 130 m to 625 m, and 15.9 m to 41 m respectively. Mean stems per hectare was 929 stems with a range from 93 to 3533 stems per hectare.

In the Southland and Westland regions, mean tree ages were 31 and 26 years, with a minimum 7 and 5 years, and a maximum of 78 and 60 years respectively. Mean site indices were 29.2 m for Southland and 30.8 m for Westland, and distributed from 19 m to 42.6 m for Southland and 11 m to 37.4 m for Westland. Altitudes ranged from 50 m to 625 m for Southland and 0 m to 330 m for Westland, with means of 217 m and 228 m respectively. Stems per hectare varied from 130 to 3366 in Southland and 173 to 4025 in Westland, with respective means of 952 and 922 stems per hectare. The relationships between stand age and site index and stand age and stocking respective to regions are plotted in Figures 3.1 and 3.2 respectively. Replications of different stocking/ha as follows: 870 plot measurements had less than 500 stems/ha, 746 plot measurements had 500 to 1000, 553 had between 1000 and 2000, and 306 had more than 2000/ha.

A summary of these data as well as mean top height, basal area/ha, and volume/ha, are represented in Table 3.2. Stem volume/ha was re-calculated using Temu's (1992) tree volume and taper equations by NZFRI in 1992. A full list of the data obtained is shown in Appendix I.

Table 3.1 Distribution of Permanent Sample Plots

Region	Number of Permanent Sample Plots	Number of re-measurements
Canterbury (CY)	57	248
Nelson (NN)	136	1247
Southland (SD)	128	666
Westland (WD)	45	314
Total	366	2475

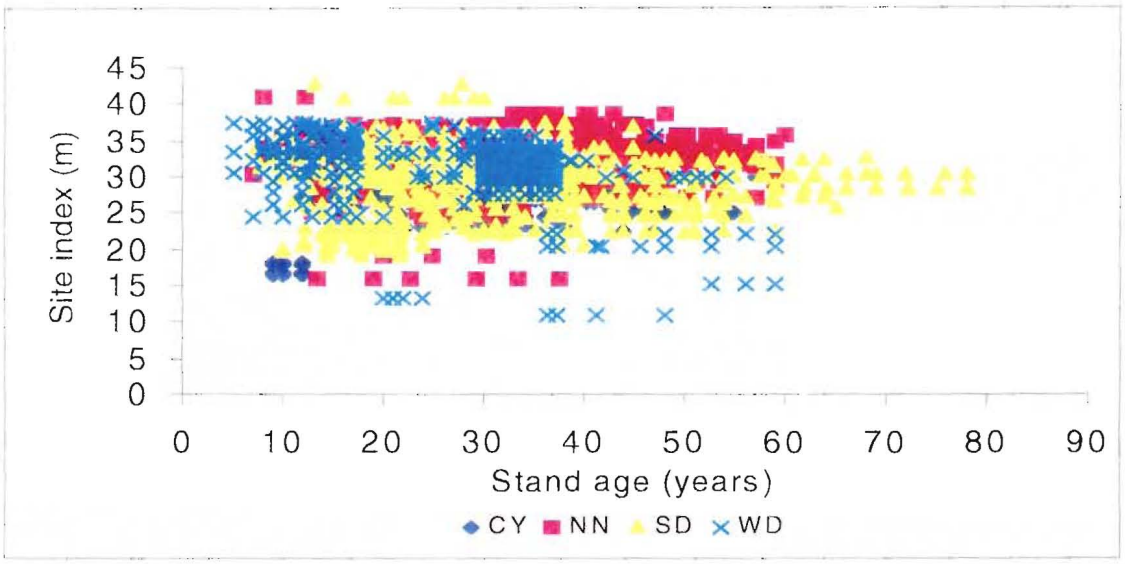

Figure 3.1 Plot of stand age vs site index by regions with PSP data.

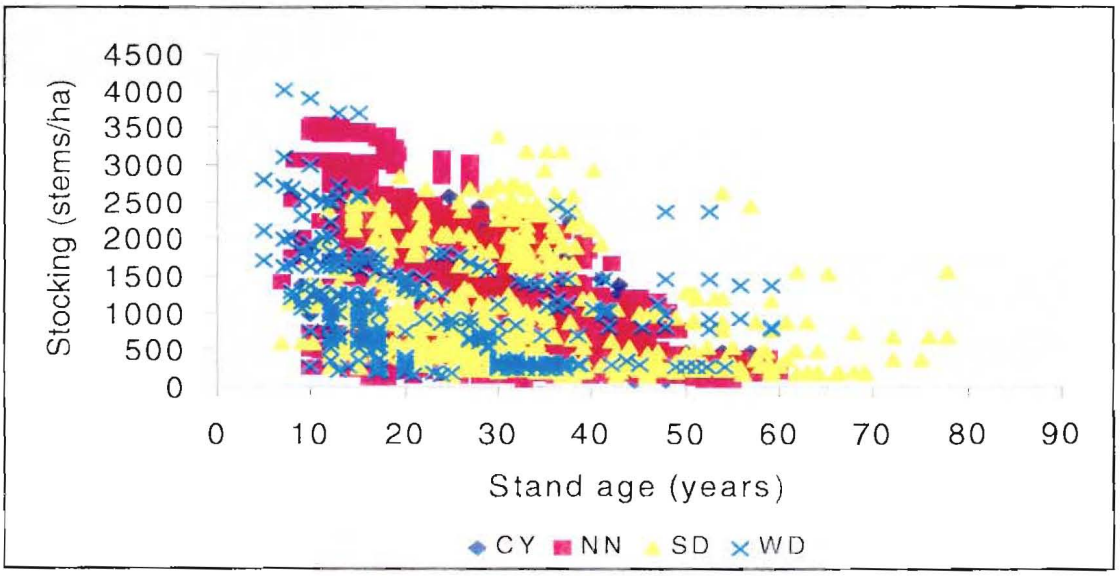

Figure 3.2 Plot of stand age vs stocking by regions with PSP data.

Table 3.2 Summary and values derived from PSP data

Region	Number of measurements	Variables	Mean	Minimum	Maximum
Canterbury	248	T (years)	28.5	9.0	57.0
		h_{100} (m)	19.8	2.9	37.1
		G (m^2/ha)	38.2	0.4	83.4
		N (stems/ha)	1033	74	2970
		V (m^3/ha)	292.9	0.6	823.3
		ALT (m)	329.5	150.0	790.0
		SI (m)	26.5	16.4	31.9
Nelson	1247	T (years)	27.9	7.0	60.0
		h_{100} (m)	22.9	5.6	45.7
		G (m^2/ha)	41.1	1.2	115.3
		N (stems/ha)	928.8	93	3533
		V (m^3/ha)	370.3	17.1	1312.1
		ALT (m)	441.3	130.0	625
		SI (m)	32.6	19.5	41.0
Southland	666	T (years)	31.0	7.0	78.0
		h_{100} (m)	21.0	5.6	46.6
		G (m^2/ha)	46.6	1.1	105.8
		N (stems/ha)	952	130	3366
		V (m^3/ha)	373.3	10.3	1118.2
		ALT (m)	217.1	50.0	625
		SI (m)	29.2	19.0	42.6
Westland	314	T (years)	25.8	5.0	59.1
		h_{100} (m)	16.6	1.9	42.0
		G (m^2/ha)	30.1	0.01	123.8
		N (stems/ha)	922	173	4025
		V (m^3/ha)	208.1	0.2	1449.7
		ALT (m)	227.7	0.0	330.0
		SI (m)	30.8	11.0	37.4
Total number of observations	2475				

The original data set was transformed into projection format. A SAS program, shown in Figure 3.3 by way of example, was used to create the projection data from the original data set.

```
DATA TEMP;
INFILE  (FILE NAME);
INPUT AGE YIELD;
AGET1=LAG (AGE);AGET2=LAG2 (AGE); · · ·
IF PLOT NE LAG(PLOT) THEN AGET1= .;
IF PLOT NE LAG2(PLOT) THEN AGET2= .;
        .
        .
        .
YIELDT1=LAG (YIELD); YIELD2=LAG2 (YIELD); · · ·
IF PLOT NE LAG(PLOT) THEN YIELD1= .;
IF PLOT NE LAG2(PLOT) THEN YIELD2= .;
        .
        .
        .
DATA TEMP1; SET TEMP;
IF AGET1 NE .THEN DO;
AGENEW1=AGET1; YIELDNEW1=YIELDT1;
PUT AGENEW1 AGE YIELDNEW1 YIELD;
END;
IF AGET2 NE . THEN DO;
AGENEW2=AGET2; YIELDNEW2=YIELDT2;
PUT AGENEW2 AGE YIELDNEW2 YIELD;
END;
```

Figure 3.3 A SAS program to create a projection data set.

The data structure for all possible growth intervals was also created with this program. If a permanent sample plot has been measured n times the possible number of combinations is $\binom{2}{n}$, although the number of combinations is actually less than $\binom{2}{n}$, because of the exclusion of combinations of re-measurements which include thinning within the projection interval.

An example of sample plot measurements and a typical permanent sample plot data format used in this study are shown in Tables 3.3 and 3.4 respectively. From this data set three projection data sets were created; 1) a short interval data set which contained measurements intervals with fewer than 3 years between T_1 and T_2; 2) a long interval data set which contained measurement intervals greater than 8 years; 3) mixed short and long intervals.

Table 3.3 Example of sample plot measurements

T (years)	G/ha	N/ha	Thinning
22	22.5	550	
24	27.8	550	
25	30.4	550	Before thinning
25	27.9	450	After thinning
26	29.9	450	Before thinning
26	26.0	400	After thinning
28	30.7	400	
32	36.3	400	

Table 3.4 Example of projection format data corresponding to re-measured data in Table 3.3

T1	T2	G1	G2	N1	N2
22	24	22.5	27.8	550	550
22	25	22.5	30.4	550	550
24	25	27.8	30.4	550	550
25	26	27.9	29.9	450	450
26	28	26.0	30.7	400	400
26	32	26.0	36.3	400	400
28	32	30.7	36.3	400	400

3.2 Quantity and Quality of the Data

It is important to verify and screen data prior to any model estimation to ensure sound data for growth and yield studies. Validation of the basic variables of mean top height (h_{100}), basal area/ha (G), stocking/ha (N) and volume/ha (V) was done by plotting these values against age for all plots to detect abnormal growth patterns. Graphical portrayal of these variables against age gave a good indication of the validity of the data: any irregularities in annual trends were detected and further analysed. Where possible obvious errors were corrected, but original records were not available so this was not always feasible. Where no corrections could be objectively made data were excluded, as in the following cases:

1) any measurement pair with a negative value of current annual increment (CAI);

2) measurements which were unreasonably high and low in a particular region for short periods over one or two years;

3) plots which contained fewer than two measurements over time, as these are not suitable for growth and yield studies; and

4) measurements where stocking/ha increased between annual measurements.

3.3 Sources of Variation

Both accuracy and precision of basic data are most important for ensuring a sound fit for growth and yield models. Fits can be improved through classifying the sources of variation in the projection equations. The following factors were further considered during the analysis of the data:

1) locality;
2) altitude;
3) thinning;
4) correlated errors through analysing repeated measures.

3.3.1 Locality

When predictions are needed for particular regions, general growth and yield models sometimes show inaccuracy and bias, and forecasted yields lose their reliability. Local adaptations expressed as local growth functions have been considered synonymous with site quality in even age stands (Temu, 1992). Several factors, such as altitude, latitude, prevailing winds, distance from the sea and topography affect site quality- soil moisture, soil development, soil type, temperature and consequently tree growth in a given the location. Some of these factors could have had a major impact on the growth of trees in this study. It was considered that for this study, dummy variables would be appropriate in representing growth trends in each region if the above environmental factors did not describe growth distinctively within each region. Dummy variables have often been found to be more effective in explaining differences in basal area, mean top height and

mortality trends than, for example, the use of site index (mean top height at age 40 years for Douglas-fir).

3.3.2 Altitude

Altitude reflects largely influences of temperature, soil fertility, rainfall and winds, which environmental factors contribute directly to the growth of trees at a given location. Thus, altitude may represent an index for all these in combination to some extent. Altitude is sometimes a useful substitute for site index, as they often have the same representation effect in New Zealand. Altitude is a more stable predictor variable than site index, which can fluctuate from one measurement to the next. Woollons and Hayward (1985) used altitude as an independent variable in a site index equation for radiata pine in the Central North Island of New Zealand. Temu (1992) also used altitude as an explanatory variable for deriving basal area/ha equation for Douglas-fir in the South Island of New Zealand. Mason (1992) used altitude to partly represent site quality for juvenile radiata pine growing in the Central North Island of New Zealand.

Altitude, which has the range of from 0 to 790 m in the data set for this study, was introduced to confirm whether or not it is a useful independent variable for projection equations. Findings are described in Chapter 4.

3.3.3 Thinning

Thinning is the process of artificially reducing the number of trees in a stand or plot. It has been one of the most important silvicultural options, because it influences growth by controlling density, spacing and distribution of trees in terms of vigour, size and quality.

Bailey and Ware (1983), for example, proposed a single basal area equation for predicting growth and forecasting yield for thinned as well as unthinned even aged stands through the use of thinning indices. Until this study materialized, separate basal area equations for thinned and unthinned stands were employed by many researchers (e.g. Pienaar and Tunbull, 1973; Piennar, 1979; Clutter and Jones, 1980; Pienaar and Shiver, 1984; and Pienaar et al., 1985).

The method of using thinning index is preferred because it does not require two independent data sets, thinned and unthinned to develop the index but instead uses average diameter before and after thinning. Murphy and Farrar (1988) made progress through developing a technique for introducing thinning into basal area projection equations as independent variables. The approach used in this study is based on that described by Murphy and Farrar (1988). Data from 191 thinned plots out of 366 plots were available for analysis in this study. A summary of thinning operations, which were applied to the sample plots, is given in Table 3.5.

Table 3.5 Main thinning regimes by regions

Region	Number of Plots Thinned					
	None	1st	2nd	3rd	Final	Total
Canterbury	32	19	6	0	0	57
Nelson	55	62	17	1	1	136
Southland	68	45	15	0	0	128
Westland	20	20	3	1	1	45
Total	175	146	41	2	2	366

3.3.4 Correlated Errors

Permanent sample plot data collected from repeated re-measurements of trees on the same plot are not uncorrelated which contravenes a requirement for classical regression. When such data are used in least-squares regression analysis, the standard error of the regression coefficients and residual errors of the fitted equation are underestimated (Sullivan and Clutter, 1972; Woollons and Hayward, 1985; Garcia, 1991). Alternative methodology has been advocated by some researchers to overcome this problem. For example, use of maximum-likelihood estimates (Sullivan and Clutter, 1972) and generalized least-squares (Ferguson and Leech, 1978; Davis and West, 1981). Most of the researchers concentrated on one particular model and paid attention to problems of hypothesis testing with their techniques. However, for analysis of growth and yield data, least-squares regression has been considered to be adequate for parameter estimation, as more emphasis is put on fitting the equations to ensure that logical relationships exist between dependent and independent variables, and to test whether or not the residuals are unbiased (Draper and Smith, 1981) rather than on precision *per se*. In this study, graphical plotting of residual and predicted or independent variables were used, together with statistical analysis and interpretation of the patterns of residuals, to avoid, to as great an extent as possible, bias in fitting equations.

3.4 Modelling Methodology

3.4.1 Creation of Data Set

The data set used in this study was obtained from NZFRI, New Zealand. Values for principal variables were extracted from PSP summaries, which show the following:

RE: region of plot location;
FO: forest of origin of data;
CP: compartment;
PLOT: plot number of identification;
SP: sub-plot;
S: site index;
ALT: altitude of plots in metres above sea level;
T_1: age in years at the beginning of a growth period:
T_2: age in years at the end of a growth period;
HT_1: mean height in metres at age T_1;
HT_2: mean height in metres at age T_2;
MTH_1: mean top height in metres at age T_1;
MTH_2; mean top height in metres at age T_2;
G_1: net basal area of the stand (m^2/ha) at age T_1;
G_2: net basal area of the stand (m^2/ha) at age T_2;
Gb: net basal area of the stand (m^2/ha) before thinning;
Ga: net basal area of the stand (m^2/ha) after thinning;
N_1: number of stems/ha at age T_1;
N_2: number of stems/ha at age T_2;
N_b: number of stems/ha before thinning;
N_a: number of stems after thinning;
V_1: net volume of the stand inside bark (m^3/ha) at age T_1;

V_2: net volume of the stand inside bark (m^3/ha) at age T_2;

Additional variables that were derived from the above database included:

Tt: age (years) of thinning;
dt: quadratic mean diameter (cm) of trees removed in thinning;
db: quadratic mean diameter (cm) of trees before thinning; and
Xt: thinning index. Xt=1 –(dt/db)

3.4.2 Data Format

When fitting a regression equation to data sets, it is often necessary to change the order of the data, file name and data format to create smaller data sets which are subsets of the master files. This was necessary in this study to enable the data to be read by the statistical package SAS, to make the processing of data easier and faster, and to use fewer computer resources.

A separate file was prepared for each variable so that different functional forms could be fitted to single variables. If new variables, which were not in the original files, were needed for fitting equations, they were created algebraically in SAS, separately for basal area/ha, mean top height and stem survival/ha. The files are presented in diskette form in Appendix II, under the following descriptions, while names of the variables used here are represented in section 3.4.1.

SHORTG.SAS, LONGG.SAS and MIXEDG.SAS.

The short (less than 3 years between measurements), long (more than 8 years between measurements) and mixed (combined previous short and long intervals) interval data and SAS programs for net basal area/ha equations contained the variables: RE, FO, CP, PLOT, T_1, T_2, G_1, G_2, ALT, S, Tt, Xt.

SHORTHT.SAS, LONGHT.SAS and MIXEDHT.SAS.

Data for short, long and mixed intervals and SAS programs for mean top height equations contained the variables: RE, FO, CP, P, T_1, T_2, MTH_1, MTH_2, ALT, S.

SHORTM.SAS, LONGM.SAS and MIXEDM.SAS.

Data and SAS programs for stem survival/ha equations contained the variables: RE, FO, CP, PLOT, T_1, T_2, N_1, N_2, S.

INTG.SAS

The independent data and SAS programs for net basal area/ha equations contained the variables: RE, FO, CP, PLOT, T_1, T_2, G_1, G_2, ALT, S.

INTHT.SAS

The independent data and SAS programs for mean top height equations contained the variables: RE, FO, CP, P, T_1, T_2, MTH_1, MTH_2, ALT, S.

INTM.SAS

The independent data and SAS programs for stem survival/ha equations contained the variables: RE, FO, CP, PLOT, T_1, T_2, N_1, N_2, ALT, S.

3.4.3 Checking Reliability of Data

The reliability of the data was further checked by analysis of residuals as well as the graphical portrayals described in section 3.2. Observations which had residuals greater than 3.5 times the random normal deviate (RND) were considered to be outliers. Such outliers indicated the need to examine corresponding variables and items of data by tracing back their history in the data set. Confirmation was not always possible, however, because the original measurement records were not available. Where the obvious causes were unknown, such measurements were categorized as suspicious but the observations were ignored in most cases because some data were collected as long as 80 years ago and it is not known what caused the errors when the measurement was taken.

Several sources of error may produce outliers in growth and yield data research: e.g. incorrect reading of measuring instruments, wrong recording, and wrong calculation of derived values. This emphasized the need for reliable measurements and conducting vigorous data checks during the initial processing of data suitability for growth and yield studies.

3.5 Method of Data Analysis

The main standard analytical procedures used in this study were non-linear least-squares regression based bonoothy on PROC NLIN in SAS. Analysis of residual patterns, through PROC UNIVARIATE was used to confirm the goodness of fit equations. When regressions were fitted to data in this study, every effort was made to ensure that:

1) the dependent and independent variables conformed to biologically and mathematically realistic relationships;

2) the functions used were of an appropriate form to represent the intended relationship;

3) a good fit was produced without bias in the regression coefficients;

as Temu (1992) and Liu Xu (1990) described in their growth and yield studies. Various sigmoid functions (see Chapter 4 for details), were fitted to the data set using PROC NLIN (SAS Inc, 1990) and the derivative-free algorithmic methods for non-linear least-squares (Ralston and Jennrich, 1979). Since residual errors were underestimated because of data correlation for repeated measures, conventional statistical analyses were inadequate and could not be accepted as valid evidence. Instead, residual analyses procedures above were used as indications of the goodness of fit of the equations and in making inferences about the statistical results.

3.5.1 Mean Square Error and Confidence Interval

Mean square error (MSE) values were used, however, to compare the relative fits of the equations applied to the various data sets. The equation with the smallest values was chosen to evaluate for further improvements through the addition of explanatory variables such as altitude, site index and dummy variables. Coefficient estimates were retained only if the coefficient estimates were apparently significant at the 5 percent significance level. Confidence intervals for each coefficient estimation were examined to ensure that the lower and upper limits were of the same sign, meaning that the confidence interval did not include zero. But no absolute evaluation was inferred.

3.5.2 Residual Patterns

Graphical charts and plots were used to check the distributions of residuals with regard to normality of errors. Residual errors were plotted against predicted values and predictor variables such as age, in order to determine goodness of fit. Whether or not the residuals patterns lay normally about the zero reference line was the main criterion for

judging the independent distribution or heterogeneous variance with the predictor variables.

Residual frequency charts illustrated the shapes of distributions, and provided valuable indications on skewness and any apparent departure from normal distributions of residuals. The SAS procedures PROC PLOT and PROC CHART were used to support analyses of residual patterns with PROC UNIVARIATE, as explained below.

3.5.3 Univariate Procedures

The PROC UNIVARIATE procedure was used to examine the residuals and provide several statistics which are valuable for making inferences about residual patterns. The univariate procedure differs from other SAS procedures which produce descriptive statistics because it provides greater detail on the distribution of a variable. PROC UNIVARIATE can provide the following:

- details on the extreme values of residuals
- quantiles, such as median
- frequency tables
- several plots to illustrate the distribution
- test of central location
- a test to determine whether residuals are normally distributed.

The important values utilized here are as follows.

1) Mean of Residuals and Absolute Mean of Residuals

The mean of residuals should be zero or close to zero in relation to the assumption that values are normally distributed with mean zero and constant variance. The absolute

mean of residuals should be as low as possible and is a measure of the average error prediction of the equations.

2) Skewness

Skewness is a measure of symmetry in that it provides inferences that can be drawn on the tendency of deviations to be larger in one direction than the other. The skewness of a normal distribution is zero, but in practice values of this lesser or greater than zero result from least-square regression: long tails to the left are indicated by negative values and long tails to the right positive values. The greater the magnitude of the value, the greater is skewness and the more unsatisfactory the fit.

3) Kurtosis

Kurtosis is the heaviness of tails in a distribution (SAS Inc, 1990) and the value of kurtosis is supposed to lie between –2 and $+\infty$. High values of kurtosis indicate that statistical methods based on the normality assumption may be inappropriate. In this study high values of kurtosis were found to be associated with the presence of outliers, which were traced back again in the data set and corresponding equations were re-assessed.

4) Extreme values

Extreme values are measures of the maximum and minimum residuals of the variable being modelled. These values listed in SAS outputs were used to identify outlier values of equation and bias. Similarity between the magnitude of positive and negative values normally indicates less bias, while dissimilar extreme positive and negative values indicated a biased residual pattern.

3.5.4 Method for Cross-Fitting

The methods of graphical analysis of residual patterns and PROC UNIVARIATE were jointly used to confirm the goodness of fit of equations for crossing fitting. Cross fitting, which examined the extent of compatibility among the predicting equations developed from the three interval data sets for state variables, involved applying each model to the other data sets and analysing the resulting residuals.

Six combinations of applying the coefficients from three relevant equations which had been found best for each interval data set were evaluated for basal area/ha, mean top height and stem survival/ha. Accordingly, the best model in the short interval data set was applied to the long and mixed interval data sets, similarly the best models in the long and mixed interval data were applied to the other two data sets. Goodness of fit of the equations was then evaluated.

3.6 Summary

In order to decide the best model, candidate models were assessed by seeking a minimum mean square error, and rigorous inspection of various residual plotting. These included plot of residuals against predicted values, as well as independent variables. Bias in prediction was also examined by normal probability plots, seeking a near zero average value, and conforming to a normal distribution. Also, the following aspects were considered:

1) the residual patterns should have no bias;

2) the residuals should have a normal frequency distribution;

3) the regression coefficient estimates of the 95 % confidence interval should have the same sign;

4) the average mean of residual and skewness should be close to 0, and kurtosis should lie between –2 and +∞;

5) extreme values of residuals between positive and negative values should be similar.

The above tests, therefore, were used together, not just on their own, to avoid biased results and provide good fits to the equations.

In terms of growth interval lengths between measurements, data were divided into 3 subsets as described in 3.1. These 3 data sets were used separately to develop stand growth and yield models as described in Chapter 4.

Chapter 4

Developing Stand Growth Models

This chapter describes the development of whole stand growth and yield model improvements for Douglas-fir grown in the South Island of New Zealand. The methodology emphasized in developing the equations here involves re-arrangement of the data to reflect different interval lengths between re-measurements. The approach also involves fitting single equations to each of three state variables, mean top height (h_{100}), basal area/ha (G) and stocking/ha (N), which is different from the approach taken by Law (1990), who followed the method advocated by Garcia (1984) of fitting the above three variables simultaneously to successive re-measurements, almost all of which were short intervals. The methodology adopted here was similar to that of Temu (1992), except that he too used only short interval data. For each possible interval length, equations were developed for (1) basal area/ha; (2) mean top height; and (3) surviving stems/ha.

Growth and yield equations were fitted to the South Island Douglas fir permanent sample plot data using various non-linear regression procedures in SAS (SAS Inc, 1990). Polymorphic and anamorphic forms of the Schumacher (Schumacher, 1939; Woollons, 1988; Clutter and Jones, 1980; Woollons and Wood, 1992), Hossfeld (Woollons et al., 1990), Chapman-Richards (Pienaar and Turnbull, 1973; Goulding,

1979) and Gompertz (Nokoe, 1978; Whyte and Woollons, 1990) were fitted, together with Weibull (Yang et al., 1978; Goulding and Shirley, 1979), monomolecular, Morgan-Mercer-Flodin and Umemura (Umemura, 1984) formulations. The location, altitude and other general features of the sample plots were also recovered. The general functional forms of equations used are presented in Table 4.1.

Table 4.1. General form of projection equations applied to data

Equation name	Equation Forms
A. Hossfeld Polymorphic	$Y_2 = 1/((1/Y_1)\,(T_1/T_2)^{\gamma} + (1/\alpha)\,(1 - (T_1/T_2)^{\gamma}))$
B. Hossfeld Anamorphic	$Y_2 = 1/((1/Y_1) + \beta(1/T_2^{\gamma} - 1/T_1^{\gamma}))$
C. Schumacher Polymorphic I	$Y_2 = \exp(\ln(Y_1)\,(T_1/T_2) + \alpha\,(1 - (T_1/T_2)))$
D.Schumacher Polymorphic II	$Y_2 = \exp(\ln(Y_1)\,(T_1/T_2)^{\beta} + \alpha\,(1 - (T_1/T_2)^{\beta}))$
E. Schumacher Anamorphic	$Y_2 = Y_1 \exp(-\beta\,(1/T_1^{\gamma} - 1/T_2^{\gamma}))$
F. Chapman-Richards Polymorphic	$Y_2 = (\alpha/\gamma)^{[1/(1-\beta)]}(1 - (1 - (\gamma/\alpha)Y_1^{(1-\beta)}) \exp(-\gamma(1-\beta)\,(T_2 - T_1))^{[1/(1-\beta)]}$
G. Chapman-Richards Anamorphic	$Y_2 = Y_1((1-\exp(-\beta T_1)) / (1-\exp(-\beta T_2)))^{\gamma}$
H. Gompertz polymorphic	$Y_2 = \exp(\ln(Y_1) \exp(-\beta\,(T_2-T_1) + \gamma\,(T_2^2-T_1^2) + \alpha\,(1- \exp(-\beta\,(T_2-T_1) +\gamma(T_2^2-T_1^2))))$
I. Weibull polymorphic	$Y2 = Y1 \exp(-\beta\,(T_2^{\gamma} - T_1^{\gamma})) + \alpha(1-\exp(-\beta(T_2^{\gamma} - T_1^{\gamma}))$
J. Weibull anamorphic	$Y_2 = Y_1(((1-\exp(-\beta T_2^{\gamma}))/(1-\exp(-\beta T_1^{\gamma})))$

4.1 SHORT INTERVAL DATA

4.1.1 Net Basal Area Projection Equation

The data set for the development of net basal area equations from the short interval data amounted to 1943 measurements, which are summarized in Table 4.2.

Table 4.2. A summary of short interval for basal area data.

Regions	N	Variables	Age (yrs)	G m^2/ha	Stems /ha	Site Index (m)	Alt (m)
Canterbury	61	Mean	28.6	36.7	799	26.91	246
		Minimum	11.8	4.4	225	23.00	150
		Maximum	46.0	73.6	2228	31.20	420
Nelson	1312	Mean	25.6	35.1	873	32.87	461
		Minimum	9.9	2.3	93	24.60	130
		Maximum	59.0	115.3	3533	38.70	625
Southland	262	Mean	30.3	41.8	820	29.17	268
		Minimum	7.0	1.1	148	20.40	50
		Maximum	78.0	99.8	3168	40.60	625
Westland	297	Mean	28.8	27.6	699	31.70	261
		Minimum	5.0	0.01	198	20.60	0
		Maximum	54.1	89.9	4025	37.40	330

Anamorphic and polymorphic forms of Schmacher, Gompertz, Hossfeld, Chapman-Richard, and Weibull equations were fitted to the data. For each of the models tested, estimation of coefficients and determination of the most applicable predictor mechanism was made using PROC NLIN in SAS.

Once the coefficients were estimated, residual values were plotted against independent variables such as age, in order to judge goodness of fit quickly. In the final choice of model, these plots of residuals were visually assessed looking mainly

for bias and numerically evaluated using the univariate procedure looking for non-normality. The coefficients of the general equations fitted to the data are presented in Table 4.3, along with the respective mean square error values.

Table 4. 3. Coefficients for general equation fitted to short interval basal area /ha data

Model Name	Coefficients/Std.			MSE
	α	β	γ	
Hossfeld Polymorphic	104.944/ 0.9402	-	2.7057/ 0.0226	1.772
Hossfeld Anamorphic	-	3.9810/ 0.5527	1.9110/ 0.0576	11.030
Schumacher Polymorphic I	5.1953/ 0.0085	-	-	1.515
Schumacher Polymorphic II	5.0802/ 0.191	1.0907/ 0.0150	-	1.487
Schumacher Anamorphic	-	19.5433/ 0.8550	0.7799/ 0.0259	4.036
Chapman-Richards Polymorphic	1.0886/ 0.0768	0.3336/ 0.0349	0.0369/ 0.0054	3.555
Chapman-Richards Anamorphic	-	0.0553/ 0.0020	2.8133/ 0.0735	4.283
Gompertz Polymorphic	4.8502/ 0.0159	0.0977/ 0.0014	0.0008 / 0.0000	1.944
Weibull Anamorphic	-	0.0014/ 0.0000	1.9098/ 0.0252	4.511

Polymorphic functional forms showed better fit than anamorphic forms of equations and most of the anamorphic functional forms were found to produce bias in their residual patterns. The Schumacher polymorphic function, equation 4.1 with MSE 1.487 was found to fit better than any of the other equations and was, therefore, selected for further analysis.

The variables used in formulating functional forms were net basal area/ha (G), age of stand (T), altitude (ALT), site index (S), thinning index (Xt), thinning age (Tt) and dummy variables representing different regions to recognize locality in the form of three dummy variables.

$$G_2=G_1(T_1/T_2)^{\beta 1}\, e^{\alpha(1-(T_1/T_2)^{\beta 1})} \qquad (4.1)$$

Numerous modifications to equation 4.1, with the addition and subtraction of predictor variables and other alterations were tested to effect further improvements. Site index (S), which is possibly the most common of all additional variables used in other studies, was also included as a predictor variable. This is defined for Douglas-fir in N.Z as the top height corresponding to the mean of the 100 largest diameter trees/ha at 40 years of age compared to the reference age of 20 years for radiata pine (Burkhart and Tennent 1977). Equation 4.2 was, therefore, fitted to the data.

$$G_2=G_1(T_1/T_2)^{\beta 1}\, e^{(\alpha+\beta 2S)(1-(T_1/T_2)^{\beta 1})} \qquad (4.2)$$

The result of including site index in the basal area equation revealed that the coefficient of site index was negative suggesting that higher site indices had lower net basal area production. This is opposite to the notion that site index and growth vary directly with one another, the basis for the long and widely accepted Eichorns's 19^{th} century law.

Altitude is sometimes substituted for site index, as they have almost the same predicting effect in New Zealand and other temperate countries. Altitude was thus introduced into the basic form, equation 4.1, instead of the site index.

$$G_2=G_1(T_1/T_2)^{\beta 1}\, e^{(\alpha+\beta 2ALT)(1-(T_1/T_2)^{\beta 1})} \qquad (4.3)$$

In order to examine the impact of locality, dummy variables (1 or 0) representing 4 separate regions, were added to equation 4.3. The Canterbury region was the default

locality. Equation 4.4 shows the form which includes three dummy variables, K1, K2 and K3, for Nelson, Southland and Westland regions respectively.

$$G_2=G_1(T_1/T_2)^{\beta 1}\, e(\alpha+\beta_2 ALT+ \beta_3 K1+ \beta_4 K2+ \beta_5 K3)(1-(T_1/T_2)^{\beta 1}) \qquad (4.4)$$

A thinning index was employed, based on that defined by Murphy and Farrar (1988), to examine the effect of thinning on basal area growth .The thinning index, Xt, employed was

$$\begin{aligned} Xt &= 1 - (Dt/Db) && \text{if } Dt/Db \neq 0 \\ &= 0 && \text{if } Dt/Db = 0 \end{aligned}$$

where

Dt = quadratic mean diameter of residual crop after thinning

Db = quadratic mean diameter before thinning.

Equation 4.5 represents the inclusion of this thinning index and the ages of any thinning in the final basal area/ha equation for the short interval data set.

$$G_2 = G_1(T_1/T_2)^{\beta 1}\exp((\alpha+ALT\beta_2+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(T_1/T_2)^{\beta 1})+\beta_6 Xt(1/T_2-1/T_1)Tt/T_2) \quad (4.5)$$

Table 4.4 presents successive improvements in which additional variables were introduced to the basic form of the Schumacher polymorphic function. All predictor variables, the coefficients for which were not different from zero at the $\alpha=0.05$ probability level were excluded and coefficients of variables that were not in line with expectations of growth relationships were excluded (e.g. when the coefficient for site index was found to be negative). It needs to be noted, of course, that classical significance testing has no absolute validity because of the mass of correlated data, and has been used simply in relative terms for ranking purposes.

Table 4.4 Successive improvements in fitting the basal area/ha equation

Input Variables	Error of Sum Squres	% Reduction in ESS	MSE
Basic form (B)	2886.5869	-	1.4872
B, Xt	2859.4404	0.94	1.4739
B,Dummy(K1, K2, K3)	2722.8053	4.78	1.4050
B, Xt, Dummy(K1,K2, K3)	2705.0101	0.65	1.3965
B, Xt, Dummy(K1,K2, K3) and ALT	2388.5237	11.70	1.2408

Improvements were evaluated mainly through graphical analysis and PROC UNIVARIATE outputs. The parameters estimated for equation 4.5 are shown in Table 4.5. The goodness of fit was evaluated through plots of residuals against predicted values, time and altitude as shown in Figures 4.1, 4.2 and 4.3. The chart of residuals is shown in Figure 4.4.

Table 4.5 Coefficients for basal area/ha equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	4.516907946	0.0520888705	2388.5237	1932	1.2408
$\beta 1$	1.113816060	0.0142512648			
$\beta 2$	0.000689254	0.000073298			
$\beta 3$	0.168941892	0.0504066814			
$\beta 4$	0.512181025	0.0497111401			
$\beta 5$	0.439476777	0.0533323672			
$\beta 6$	-4.797716026	1.6191963739			

There were no apparent biases in the short interval basal area projection with equation 4.5. The mean average residual, skewness and kurtosis values were –0.0094 m^2/ha, 0.092 and 1.204 respectively. Equation 4.5 was able to contain residuals all within ± 4.0 m^2/ha. This equation had a maximum residual of 3.92 m^2/ha, a minimum residual of –4.0 m^2/ha, while 95% of residuals were contained within ±1.92 m^2/ha.

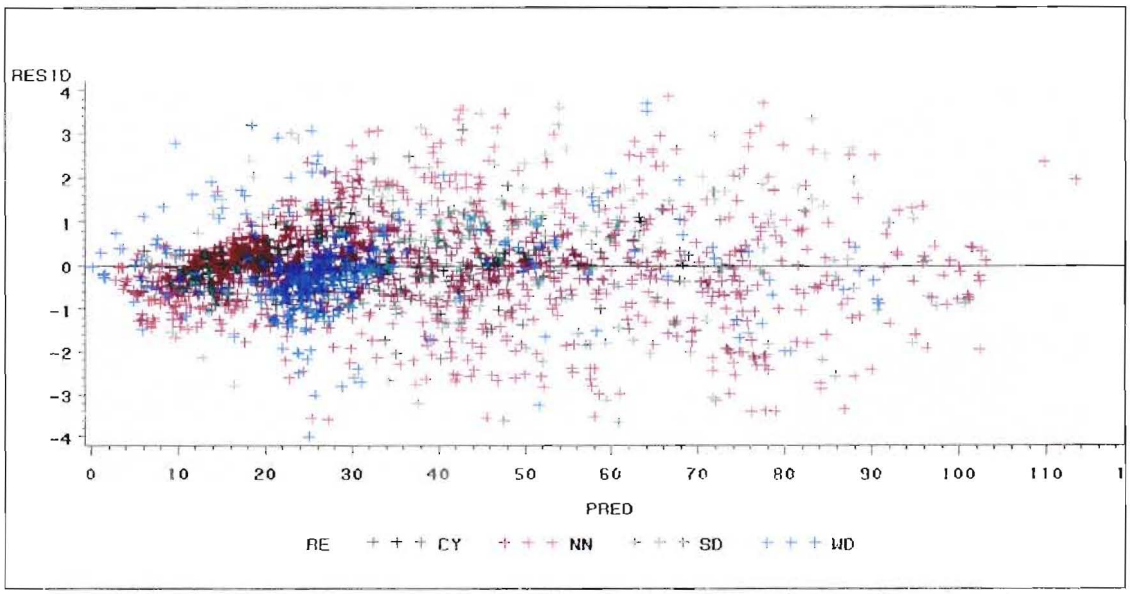

Figure 4.1: Plot of residuals vs predicted values for short interval net basal area/ha equation

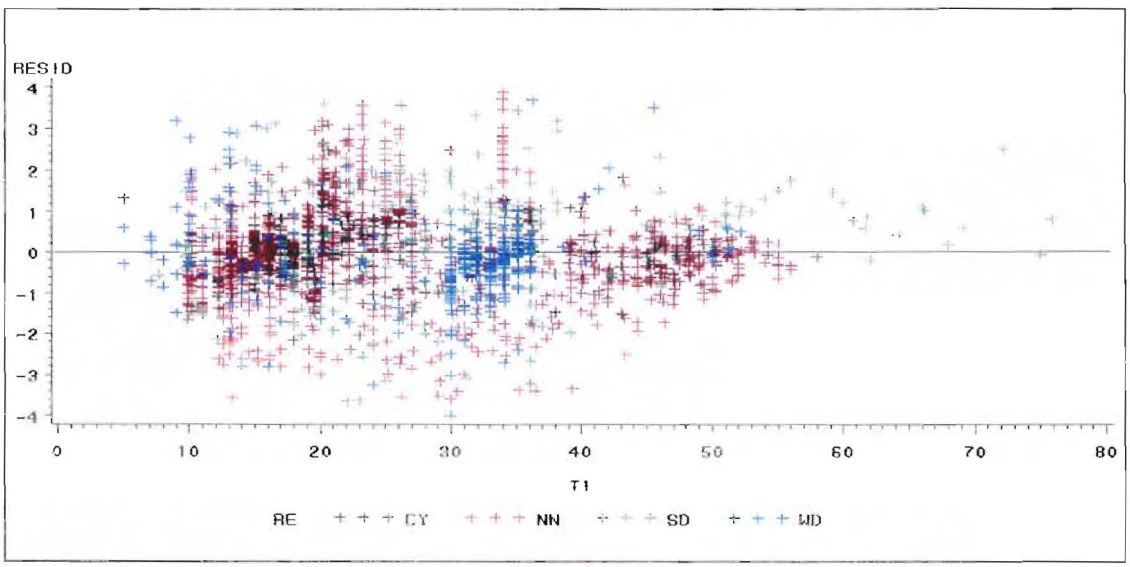

Figure 4.2: Plot of residuals vs age (T1 years) for short interval net basal area/ha equation

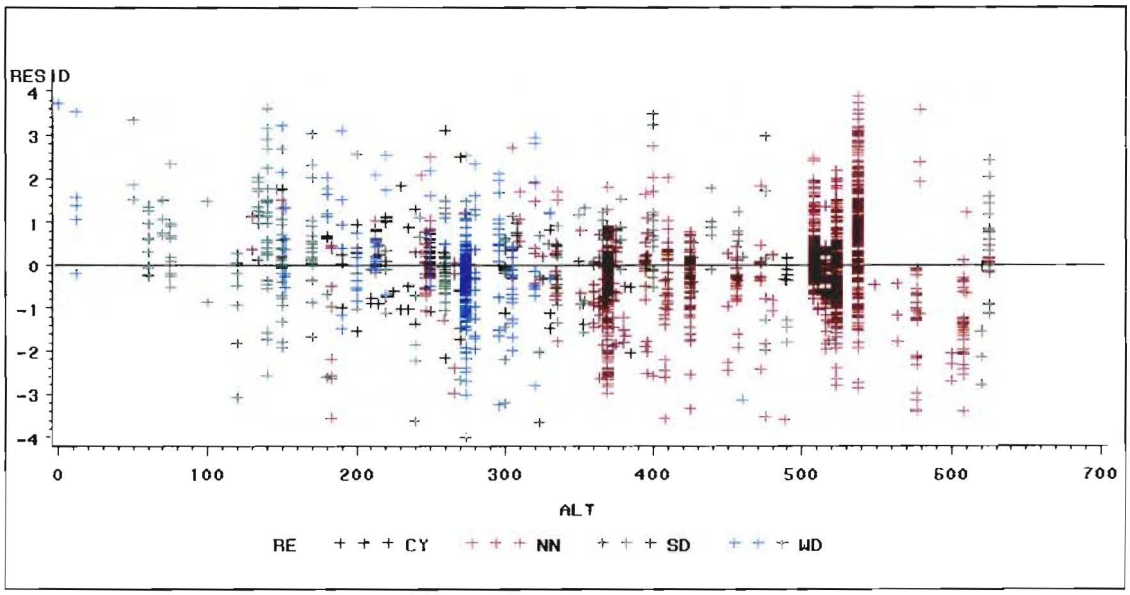

Figure 4.3: Plot of residuals vs altitude for short interval net basal area/ha equation

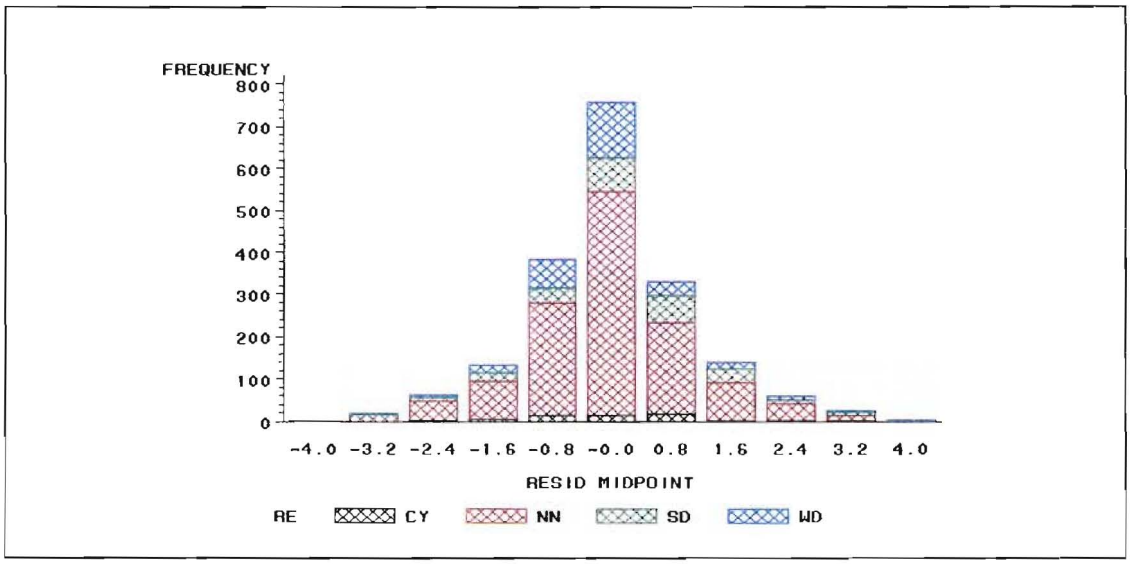

Figure 4.4: Frequency distribution of residuals for short interval net basal area/ha equation

4.1.2 Mean Top Height Equation

Mean top height represents the average height of the 100 trees of largest diameter at breast height per hectare and is often referred to as dominant height. The data which were used to develop a general mean top height equation are presented in summary in Table 4.6.

Table 4.6 Summary of data for short interval mean top height equation

Region	No.	Variable	Mean	Minimum	Maximum
Canterbury	61	Age(yrs)	28.4	11.8	45.2
		h_{100} (m)	20.1	8.3	34.6
Nelson	1295	Age(yrs)	25.4	9.9	59.0
		h_{100}(m)	21.2	7.1	45.0
Southland	253	Age(yrs)	29.8	7.0	78.0
		h_{100} (m)	21.4	4.3	45.8
Westland	296	Age(yrs)	28.8	5.0	54.1
		h_{100} (m)	22.1	1.9	37.5

Various functional forms having good possibility for fitting mean top height to the data set were tested. The equation giving the overall best fit was identified through examination of residual patterns and mean square error (MSE) values. The coefficients and the residual mean squares for candidate equations are presented in Table 4. 7.

Table 4. 7 Coefficients for general equation fitted to short interval mean top height data

Model Name	Coefficients/Std.			MSE
	α	β	γ	
Hossfeld Polymorphic	70.2989/ 2.0364	-	2.0364/ 0.0194	0.328
Hossfeld Anamorphic	-	2.7862/ 0.1147	1.3640/ 0.0181	0.459
Schumacher Polymorphic I	3.9895/ 0.0097	-	-	0.510
Schumacher Polymorphic II	5.3455/ 0.0950	0.4309/ 0.0168	-	0.320
Schumacher Anamorphic	-	8.8446/ 0.0834	0.3814/ 0.0166	0.341
Chapman-Richards Polymorphic	0.8614/ 0.0767	0.1559/ 0.0497	0.0250/ 0.0040	0.349
Chapman-Richards Anamorphic	-	0.0286/ 0.0012	1.5025/ 0.0253	0.350
Gompertz Polymorphic	4.0562/ 0.0557	0.0595/ 0.0014	0.0003 / 0.0000	0.375
Weibull Anamorphic	-	0.0065/ 0.0002	1.3501/ 0.0147	0.357

The Schumacher polymorphic equation 4.6, which had the lowest mean square error and mean residual error, was found to give the best fit to the data set.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta 1} e^{\alpha(1-(T_1/T_2)^{\beta 1})} \qquad (4.6)$$

Other predictor variables representing site specific factors were included to equation 4.6. Altitude was introduced because it has been found be an important variable for explaining variation in mean top height growth as shown in, for example Woollons and Hayward (1985); Whyte et al. (1992); Mason (1992). The result of including altitude showed that the asymptote of mean top height decreased with altitude, which is a biologically realistic trend.

Dummy variables representing regions were used to localize equation 4.6. Only one dummy variable for the Nelson region was found to be necessary, as shown in equation 4.7, because three regions (Canterbury, Southland and Westland) showed similar height growth patterns. The variables used in this equation found best were stand age (T), mean top height (h_{100}), altitude (ALT) and the dummy variable K1 for the Nelson region.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+\beta 2ALT+\beta 1K1)(1-(T_1/T_2)^{\beta}) \quad (4.7)$$

Parameter estimation for equation 4.7 is summarized in Table 4.8. All the parameter estimates were judged significant at the 5 % level.

Table 4.8 Coefficients for mean top height equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	5.334325040	0.10779474537	562.1749	1825	0.3087
β	0.433920295	0.01657165934			
$\beta 1$	0.276508010	0.05374944052			
$\beta 2$	-0.000517740	0.00018387696			

Figure 4.5 shows the plot of residuals against predicted values, and Figure 4.6 shows the plot of residuals against age. Figure 4.7 shows the plot of residuals against altitude and Figure 4.8 shows a chart of the frequency distribution of the residuals. Equation 4.7 produced residuals which had a mean residual error of 0.03, skewness of 0.35 and kurtosis of 0.59, with a maximum residual of 1.94 m, a minimum residual of –1.81 m, while 95% of residuals lay within ±1.0 metres.

4.1.3 Site Index Equation

Site index is used as a measure of potential crop productivity and is defined as the mean top height of a stand at a specified index age. Site index equations can be derived from mean top height equation 4.7, by setting T_2 equal to 40 years, which is the base age for Douglas-fir in New Zealand (Burkhart and Tennent, 1977; Mountfort, 1978).

$$S=h_{100,1}(T_1/40)^{\beta}e(\alpha+\beta 2ALT+\beta 1K1)(1-(T_1/40)^{\beta}) \qquad (4.8)$$

Alternatively, where site index is known, mean top height can be estimated for a specified age. Substituting S with required site index values results in polymorphic growth curves as shown in Figure 4.9 which used actual data.

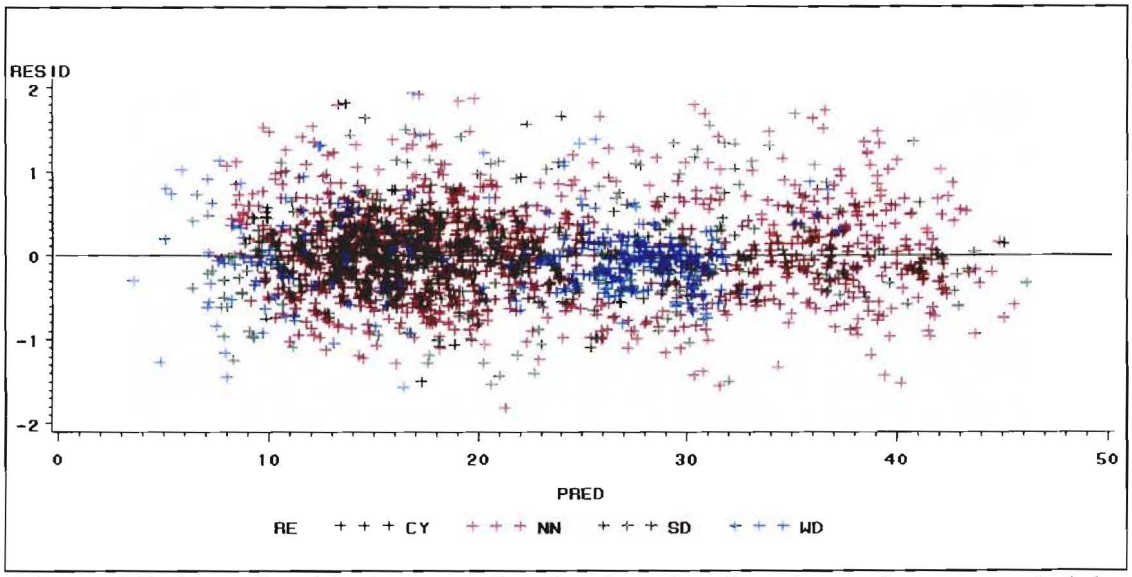

Figure 4.5: Plot of residuals vs predicted values for short interval mean top height equation

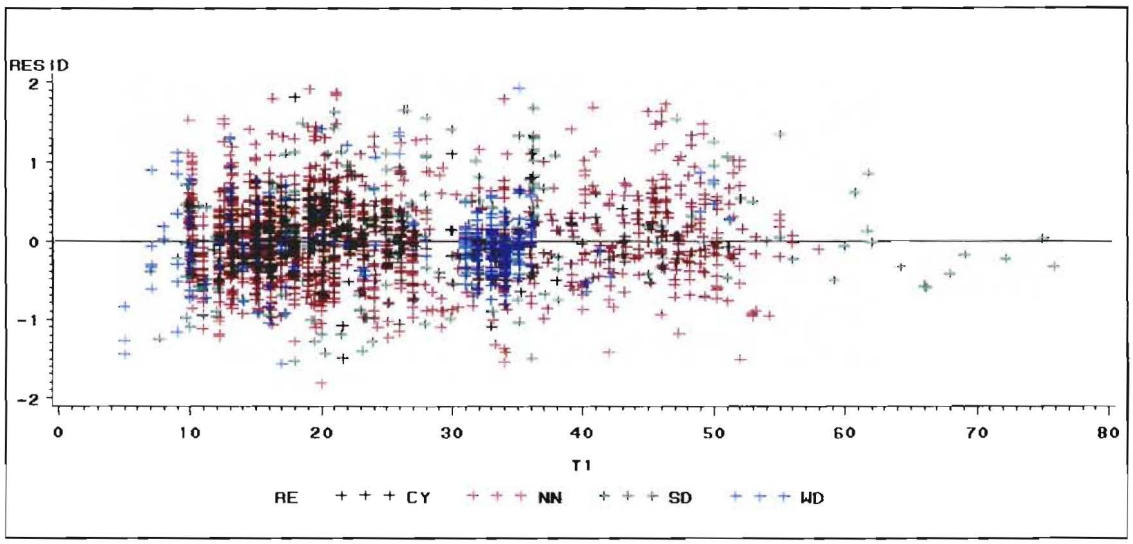

Figure 4.6: Plot of residuals vs age (T1 years) for short interval mean top height equation

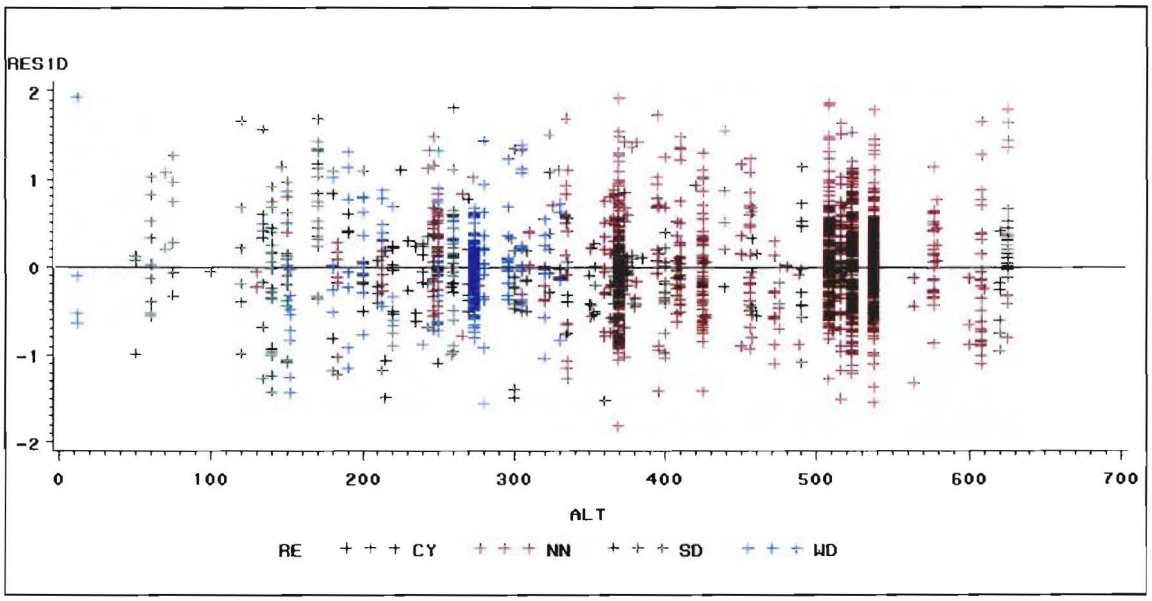

Figure 4.7: Plot of residuals vs altitude for short interval mean top height equation

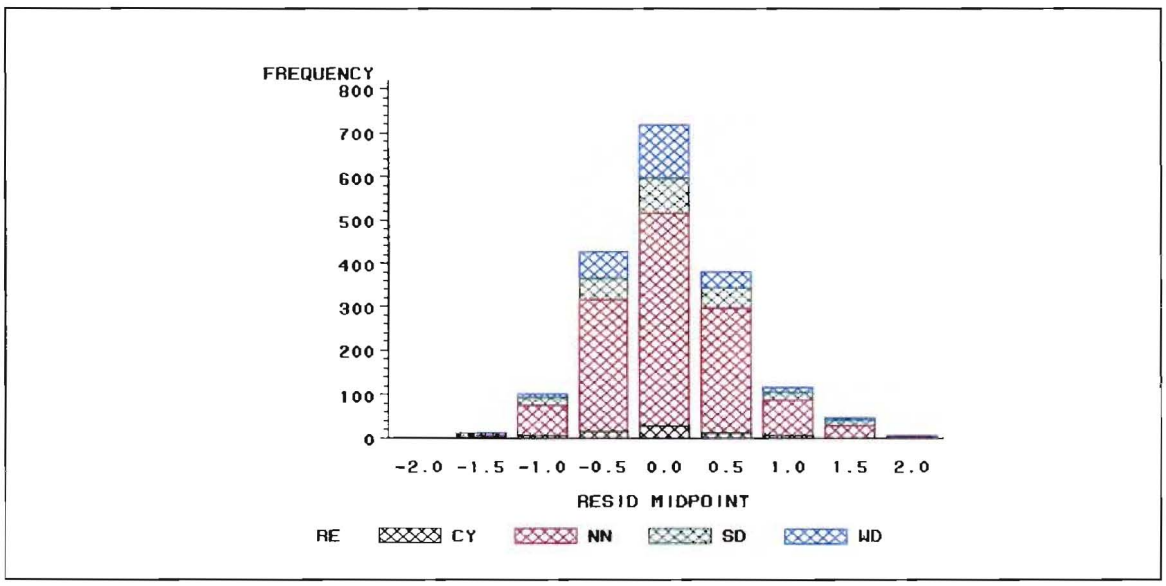

Figure 4.8: Frequency distribution of residuals for short interval mean top height equation

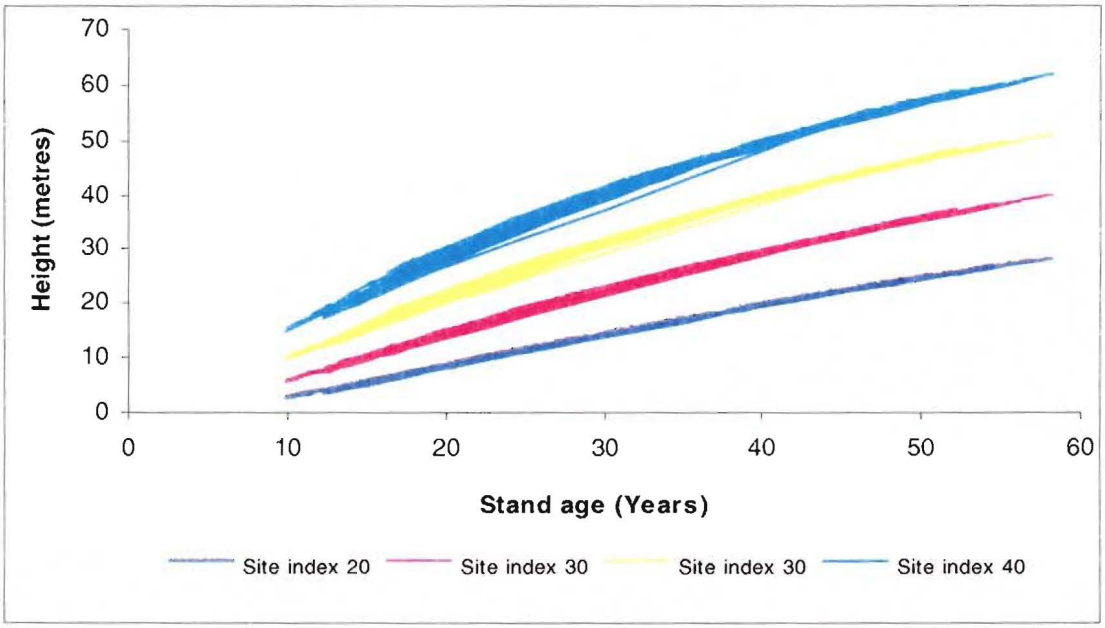

Figure 4.9: Site index curves derived from short interval data for the in Nelson region

4.1.4 Stem Survival/ha Equation

Mortality has been found almost always to be a difficult variable to predict successfully in growth models for intensively managed plantations. Mortality is often considered to be the main source of the variation that is usually evident between predicted and actual yields (Vanclay, 1994).

Re-measurement data of the type just described are generally used to fit some kind of a difference equation model that predicts N_2 as a function of T_1, T_2 and N_1. The kind of model needed should possess the logical property that for even-aged stands, if T_2 is greater than T_1, N_2 should be less than N_1, there being no ingrowth. Stem survival/ha data in this study were screened to ensure that only those intervals in which mortality occurred were included. If all the data, where no mortality has occurred, are retained, then difficulties in modelling can be encountered, because convergence is sometimes being difficult to achieve. The act of including data contrary to the model being fitted can frequently exacerbate this. Also, the goodness-of-fit of any model is largely assessed by rigorous inspection of residual patterns. But, the act of including all the data produces ill-fitting diagrams, with a substantial amount of data exhibiting bias in prediction. Subsequently, the overall performance of any candidate model cannot be evaluated easily (Woollons, 1998). The data used to build the mortality equation are summarized in Table 4.9.

Table 4.9 Summary data for short interval stem survival/ha equation

Region	No.	Variable	Mean	Minimum	Maximum
Canterbury	16	Age(yrs) Stems/ha	32.0 1159	20.6 225	45.2 2228
Nelson	301	Age(yrs) Stem/ha	27.7 1640	9.9 161	49.3 533
Southland	53	Age(yrs) Stem/ha	27.2 1322	7.7 210	54 2700
Westland	76	Age(yrs) Stem/ha	21.9 1407	5.0 222	41.7 4025

The model found to describe stem survival/ha best in the short interval data set was a modified form of the exponential equation, as shown in equation 4.9.

$$N_2 = N_1 \exp(\alpha+\gamma K1)\ (T_2^{\beta} - T_1^{\beta})) \qquad (4.9)$$

Including site index did not improve the fit. A dummy variable for regions was necessary only for Nelson. Table 4.10 sets out the parameters for the stem survival/ha equation. The plots of residuals against predicted value, age and the frequency distribution chart are given in Figures 4.10, 4.11 and 4.12. Because of a lack of observations in Canterbury, Southland and Westland, residuals for these regions show slightly imbalanced patterns.

Table 4.10 Coefficients for short interval stem survival/ha equation

Parameter	Estimates	Std. Error	ESS	N	MSE
α	-0.000979419	0.00028267916	613899.5	446	1385.8
β	1.788659236	0.07811574737			
γ	0.000336821	0.00010416432			

As in other studies, a successful fit for the stem survival/ha equation for this data set was found to be difficult. The residuals, however, lie mainly within ±55 trees per ha as shown in Figure 4.10, and the mean residual was quite satisfactory being less than 1 stem/ha. The equation 4.9 had a value of –0.73 for skewness and 1.38 for kurtosis.

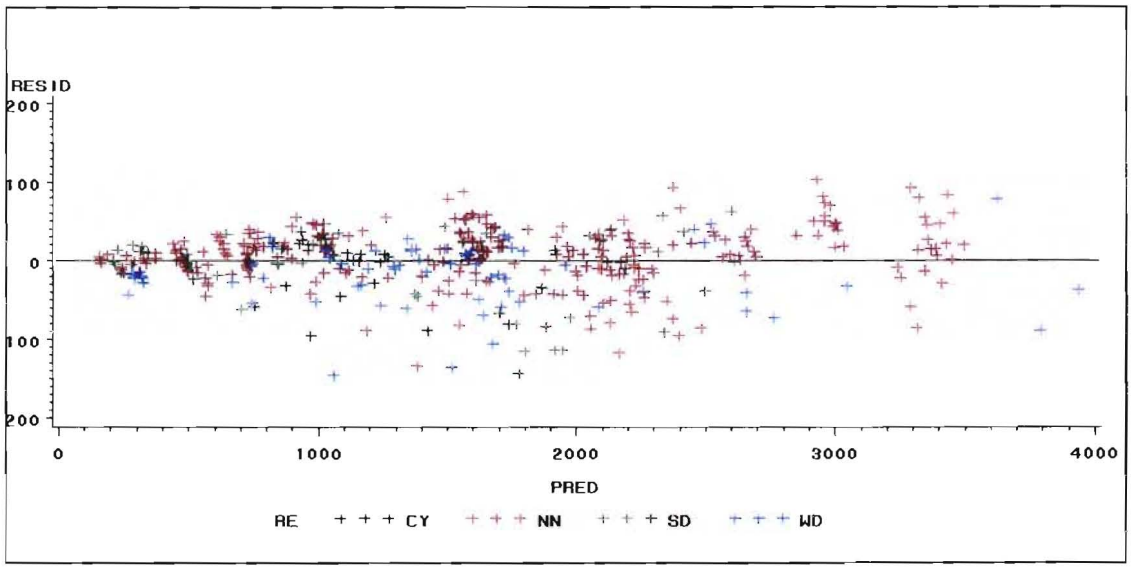

Figure 4.10: Plot of residuals vs predicted values for short interval stem survival/ha equation

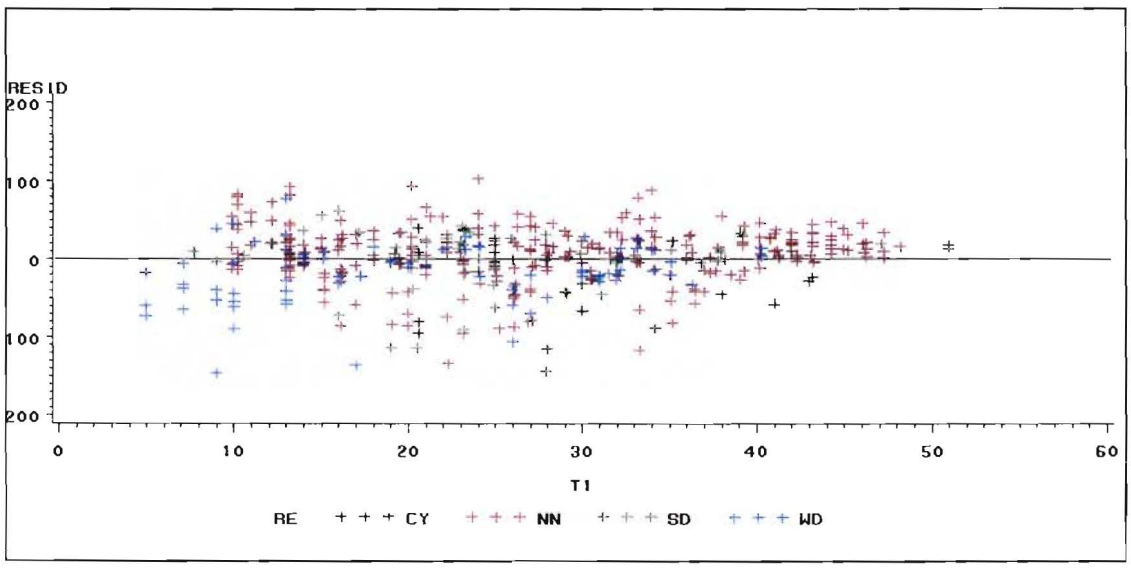

Figure 4.11: Plot of residuals vs age (T1 years) for short interval stem survival/ha equation

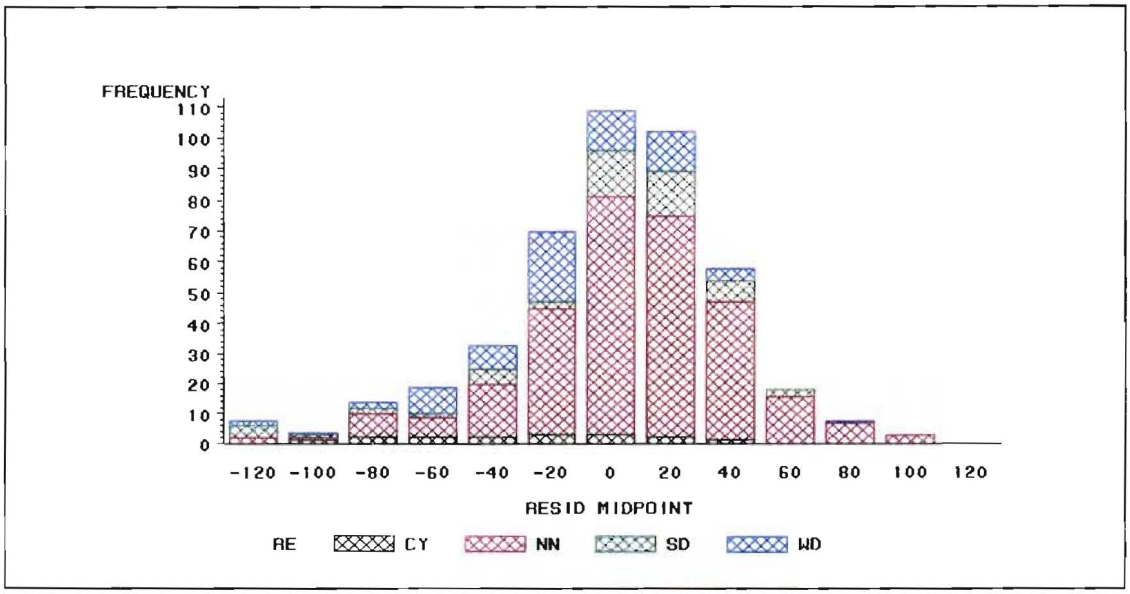

Figure 4.12: Frequency distribution of residuals for short interval stem survival/ha equation

4.2 LONG INTERVAL DATA

4.2.1 Net Basal Area/ha Projection Equation

A long interval data set was created with only those measured pairs, which contained an interval of more than 8 years were retained, an entirely different strategy to that usually employed. The long interval data contained intervals from 8 to 28 years. Table 4.11 presents a summary of the data which were used to develop a basal area equation for the long interval data set.

Table 4.11 A summary of long interval data for basal area equation

Region	N	Variables	Age (yrs)	G m^2/ha	Stems /ha	Site Index (m)	Alt (m)
Canterbury	107	Mean	25.9	42.6	1366	27.57	277
		Minimum	14.0	10.2	300	23.50	180
		Maximum	44.1	83.4	2970	31.20	470
Nelson	1202	Mean	26.3	46.3	1061	33.04	470
		Minimum	7.0	1.2	128	24.60	183
		Maximum	59.1	115.3	3533	37.30	625
Southland	391	Mean	32.2	53.6	973	30.14	202
		Minimum	7.7	4.0	148	22.00	60
		Maximum	78.0	105.8	3366	40.60	490
Westland	80	Mean	32.7	54.5	894	31.21	200
		Minimum	8.0	0.2	222	20.60	0
		Maximum	59.1	123.8	1827	37.40	320

Growth functions were fitted to the data and the equation which best fitted was chosen for additional enhancement by incorporating several explanatory variables in a logical manner. The coefficients of the general functional forms fitted to the data are presented in Table 4.12, with the respective mean square error values.

Table 4. 12 Coefficients for general equation fitted to long interval basal area /ha data

Model Name	Coefficients/Std.			MSE
	α	β	γ	
Hossfeld Polymorphic	100.5816/ 0.7856	-	2.8319/ 0.0236	60.819
Hossfeld Anamorphic	-	1.3834/ 0.2552	1.5902/ 0.0758	568.383
Schumacher Polymorphic I	5.1535/ 0.0073	-	-	64.054
Schumacher Polymorphic II	4.9499/ 0.0159	1.2050/ 0.0178	-	57.899
Schumacher Anamorphic	-	16.8092/ 1.1391	0.7857/ 0.0380	252.049
Chapman-Richards Polymorphic	0.8512/ 0.0651	0.5180/ 0.0466	0.0849 / 0.0142	68.121
Chapman-Richards Anamorphic	-	0.0525/ 0.0027	2.784/ 0.0871	259.954
Gompertz Polymorphic	4.7836/ 0.0114	0.0907/ 0.0013	0.0006/ 0.0000	50.693
Weibull Polymorphic	896.351/ 307.08	0.0199/ 0.0059	0.5605/ 0.0288	64.533
Weibull Anamorphic	-	0.0035/ 0.0002	1.6603/ 0.0326	266.611

The Gompertz polymorphic form which is shown as equation 4.10 was in this instance found to provide the best fit. Therefore, this equation was selected for further analysis.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{\alpha (1 - \exp (-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (4.10)$$

Site index was introduced to equation 4.10 to examine the potential for further improvement of the basic equation and was deemed significant at the α=0.05 level. It gave some improvement in the mean square error, but the coefficient of site index was negative which means that better sites have lower basal area/ha values. That notion is not biologically acceptable for a relationship between site index and growth. But it can be explained through height growth trends differing in one region (Canterbury) with its unusual rainfall gradient.

Site index was replaced with altitude, which improved in the mean square error and consistency among regions. Introducing a thinning index did not appear to improve the fit to equation 4.10, and it was not significant. Dummy variables were added to equation 4.10 to account for possible different growth patterns across regions. The analysis showed that each of the four regions had substantially different asymptotes. Several additions and removals of variables to equation 4.10 resulted in finding that equation 4.11, which included three dummy variables and altitude, proved best. Table 4.13 presents the successive improvement in which additional variables were introduced to the basic form of the Gompertz polymorphic function, where G, T and ALT are basal area/ha, stand age, and altitude respectively, and K1, K2 and K3 are dummy variables accounting for locations as before.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (4.11)$$

Table 4.13 Successive improvement in fitting for long interval net basal area/ha equation

Input Variables	Error SS	% Reduction in ESS	MSE
Basic form (B)	90081.15	-	50.693
B, ALT	89709.53	0.41	50.512
B,Dummy(K1, K2, K3)	77199.2781	13.95	43.5171
B, ALT, Dummy	76893.8825	0.40	43.3694

The plot of residuals against predicted values is shown in Figure 4.13, the plots of residuals against age and altitude in Figures 4.14 and 4.15 respectively. The frequency distribution chart is shown in Figure 4.16. The plots do not show any obvious bias, but there was underestimation for ages around and over 50 years in the Southland region. It was later confirmed that values of underestimation at over 60 years of age were from two plots. The parameters for equation 4.11 are given in Table 4.14.

Table 4.14 Coefficients for long interval basal area equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	4.476184855	0.02585221661	76893.8825	1780	43.3694
β	0.093867541	0.00123944678			
γ	0.000594031	0.00001751370			
$\beta 1$	0.223119181	0.2397553246			
$\beta 2$	0.348222767	0.02337645789			
$\beta 3$	0.417061618	0.02969846465			
$\beta 4$	0.000119958	0.00004552667			

The mean average residual for equation 4.11 was –0.0106 m^2/ha, confirming that slight overestimation shown graphically. Skewness and kurtosis values were 0.065 and 0.173 respectively. This equation had a maximum residual of 22.03 m^2/ha, a minimum residual of –21.66 m^2/ha, while 95% of residuals were contained within ±11.36 m^2/ha. It was found for the long interval data net basal area/ha equation that, when the length of the projection interval increased, precision decreased but without a corresponding decline in accuracy.

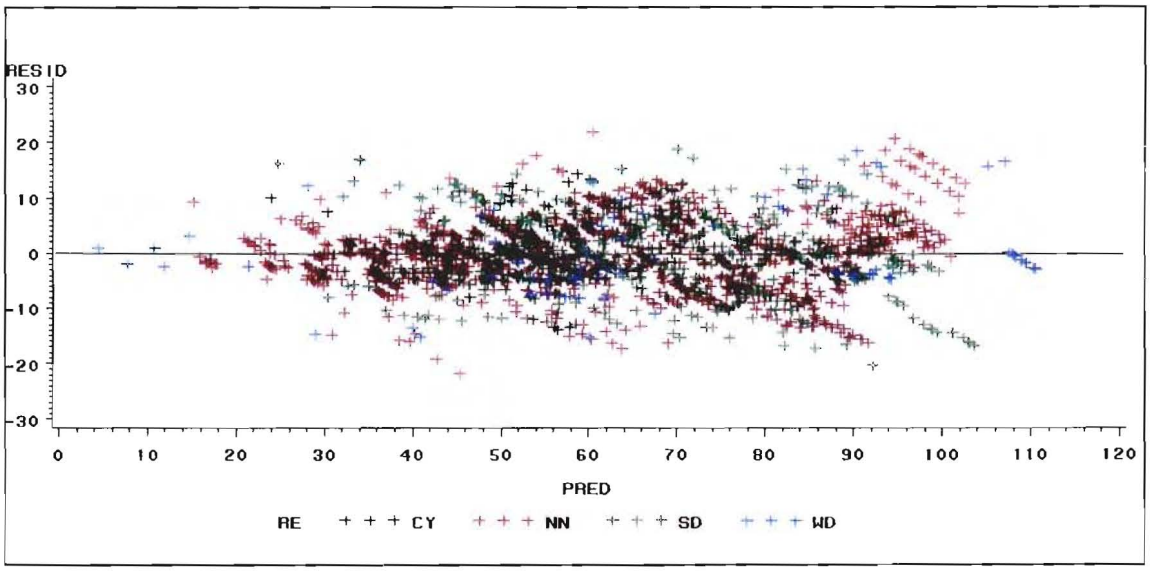

Figure 4.13: Plot of residuals vs predicted values for long interval basal area/ha equation

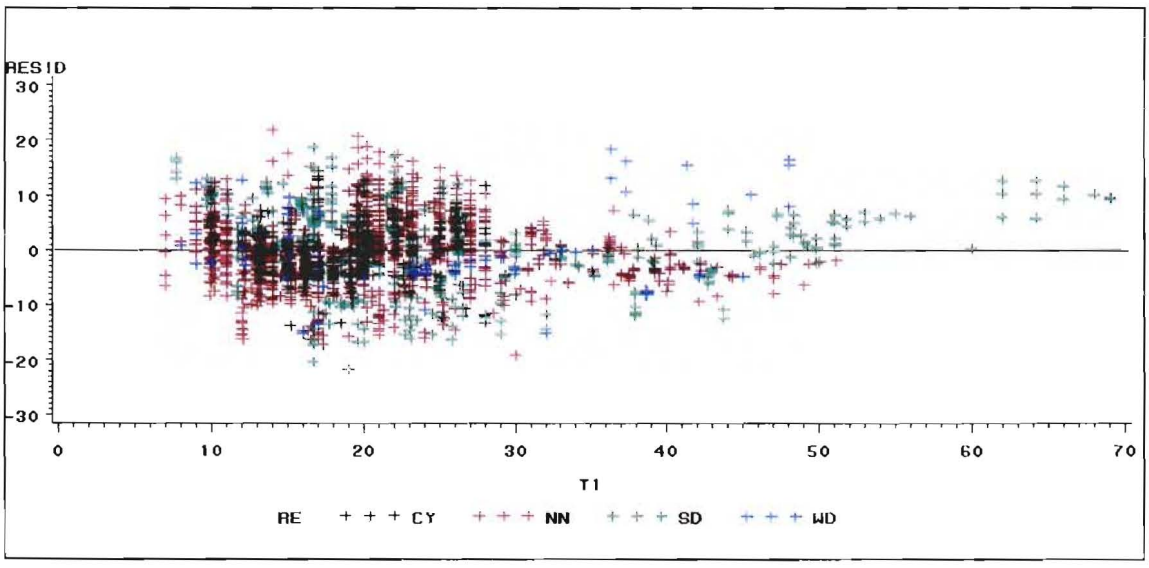

Figure 4.14: Plot of residuals vs age (T1 yerars) for long interval basal area/ha equation

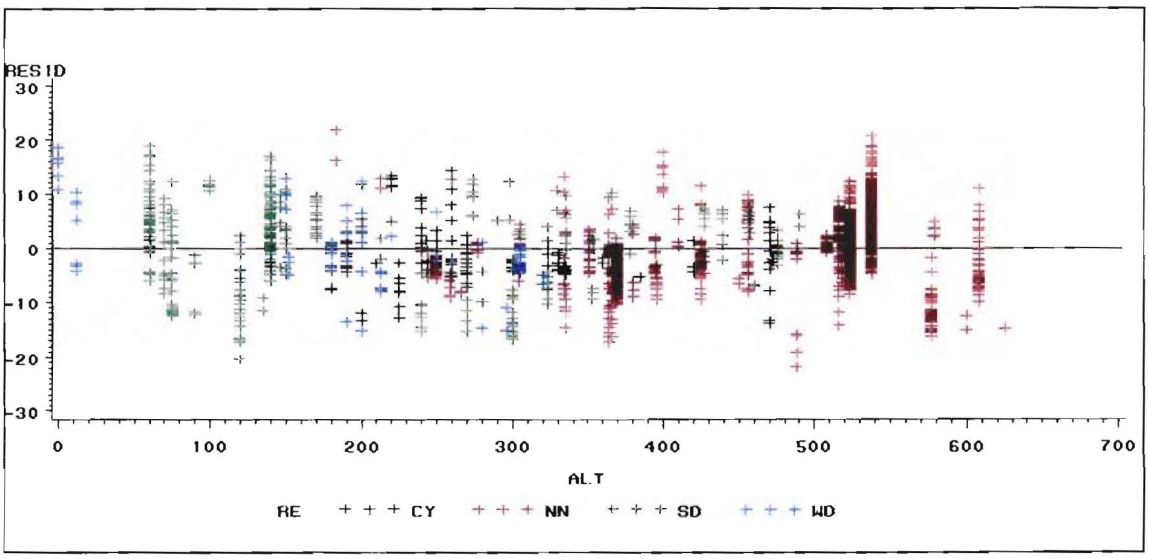

Figure 4.15: Plot of residuals vs altitude for long interval basal area/ha equation

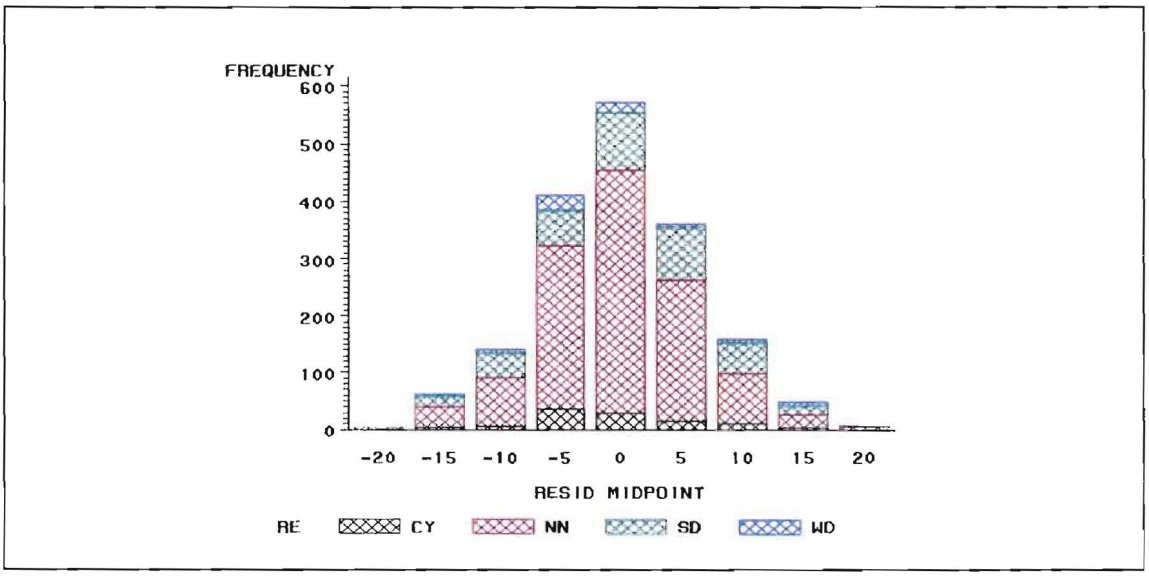

Figure 4.16: Frequency distribution of residuals for long interval basal area/ha equation

4.2.2 Mean Top Height Equation

The data used for analysing mean top height in the long interval data set are summarized in Table 4.15.

Table 4.15 Summary of data for long interval mean top height equation

Region	No.	Variable	Mean	Minimum	Maximum
Canterbury	106	Age(yrs)	26.1	14.0	45.2
		h_{100} (m)	19.1	9.7	34.6
Nelson	1201	Age(yrs)	26.3	7.0	59.1
		h_{100}(m)	22.4	5.6	45.0
Southland	390	Age(yrs)	32.2	7.7	78.0
		h_{100} (m)	23.4	5.7	45.8
Westland	86	Age(yrs)	33.8	8.0	59.1
		h_{100} (m)	23.7	3.0	42.1

Various growth equation forms were again fitted to the data set and their residual patterns compared. The coefficients of the general equations fitted to the data are presented in Table 4.16, along with the respective mean square error values.

Table 4.16 Coefficients for general equation fitted to the long interval mean top height data

Model Name	Coefficients/Std.			MSE
	α	β	γ	
Hossfeld Polymorphic	80.4825/ 1.6261	-	1.4162/ 0.0105	1.943
Hossfeld Anamorphic	-	2.0564/ 0.0606	1.2466/ 0.0133	8.965
Schumacher Polymorphic I	4.0626 0.0050	-	-	4.952
Schumacher Polymorphic II	5.8543/ 0.0791	0.3655/ 0.0101	-	1.939
Schumacher Anamorphic	-	8.9390/ 0.0340	0.2927/ 0.0090	2.403
Chapman-Richards Polymorphic	0.7043/ 0.0741	0.2855/ 0.0613	0.0376/ 0.0065	2.680
Chapman-Richards Anamorphic	-	0.0213/ 0.0006	1.3720/ 0.0116	2.395
Gompertz Polymorphic	4.1922/ 0.0324	0.0536/ 0.0010	0.0003 / 0.0000	2.365
Weibull Anamorphic	-	0.0064/ 0.0000	1.2845/ 0.0072	2.423

The equation showing best fit was chosen for further analysis to investigate improvement through the addition of site specific factors. The Schumacher polymorphic form as shown in equation 4.12 had the lowest mean square error (MSE) value and a satisfactory distribution of residuals without apparent bias. Therefore, this equation was chosen for further examination.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta 1}e^{\alpha(1-(T_1/T_2)^{\beta 1})} \quad (4.12)$$

Further analysis to include altitude showed improvement in the model but not to a great extent when regional dummy variables were introduced, as shown in equation

4.13. Evidently, height growth trends in the long interval data differed slightly from those for the short interval data, in that three dummy variables were necessary to represent different localities. Accordingly, equation 4.13 was found to provide the best fit for mean top height in the long interval data set.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/T_2)^{\beta}) \quad (4.13)$$

Coefficient estimation for equation 4.13 is provided in Table 4.17. Residual plots and a frequency distribution chart are presented in Figures 4.17, 4.18 and 4.19. The residual patterns were generally uniformly distributed without any serious bias.

Table 4.17 Coefficients for long interval mean top height equation

Parameter	Estimates	Std. Error	ESS	N	MSE
α	5.478561902	0.07319882630	2869.4550	1696	1.6969
β	0.370768578	0.00952615107			
$\beta 1$	0.378216102	0.03396904170			
$\beta 2$	0.205150963	0.03638424606			
$\beta 3$	0.437769742	0.04957317536			

PROC UNIVARIATE in SAS showed that residual statistics were quite satisfactory, as it showed that it had 0.21 value for skewness and 0.63 value for kurtosis. The equation gave a maximum residual of 4.6 m, a minimum residual of –4.5 m, a mean residual of 0.023 m and 95% of residuals lay within ±2.25 m.

4.2.3 Site Index Equation

Mean top height equation 4.13 was used to make a site index equation, 4.14, by calculating $h_{100,2}$ at age 40 years.

$$S=h_{100,1}(T_1/40)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/40)^{\beta}) \quad (4.14)$$

Site index curves for each region could be developed using equation 4.14 by substituting the S and T_2 with desired values. An illustration of the height growth trends for the four different regions represented is given in Figure 4.20.

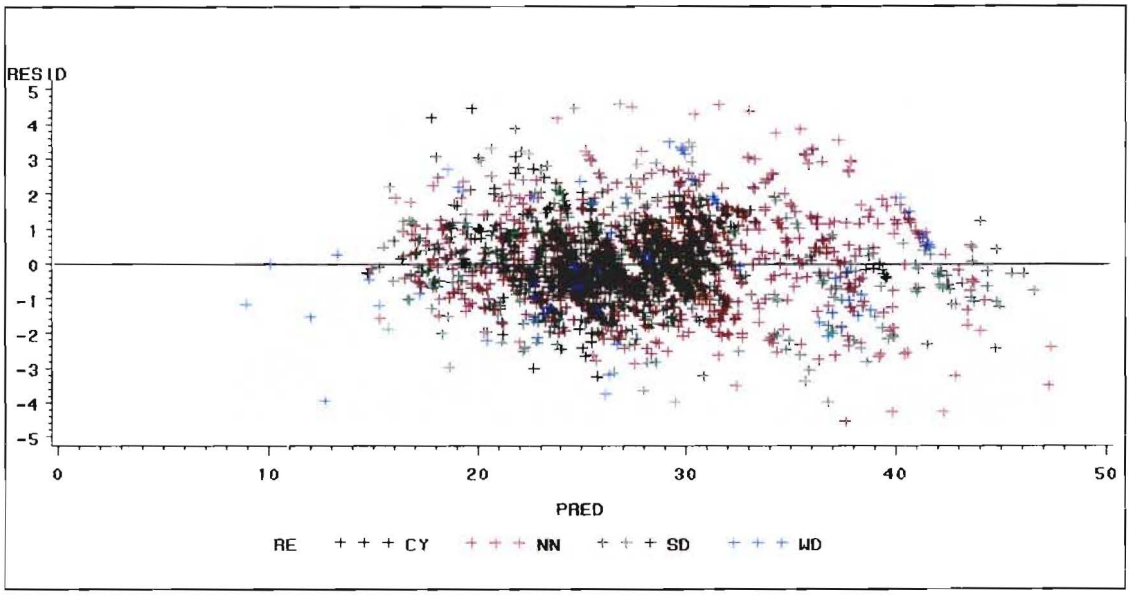

Figure 4.17: Plot of residuals vs predicted values for the long interval mean top height equation

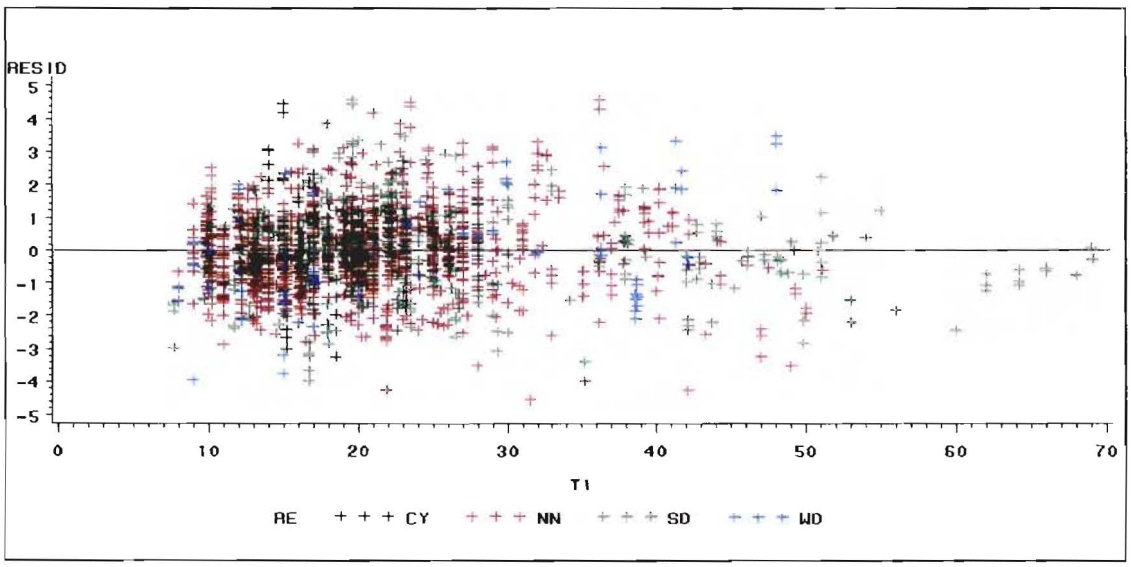

Figure 4.18: Plot of residuals vs age (T1 years) for the long interval mean top height equation

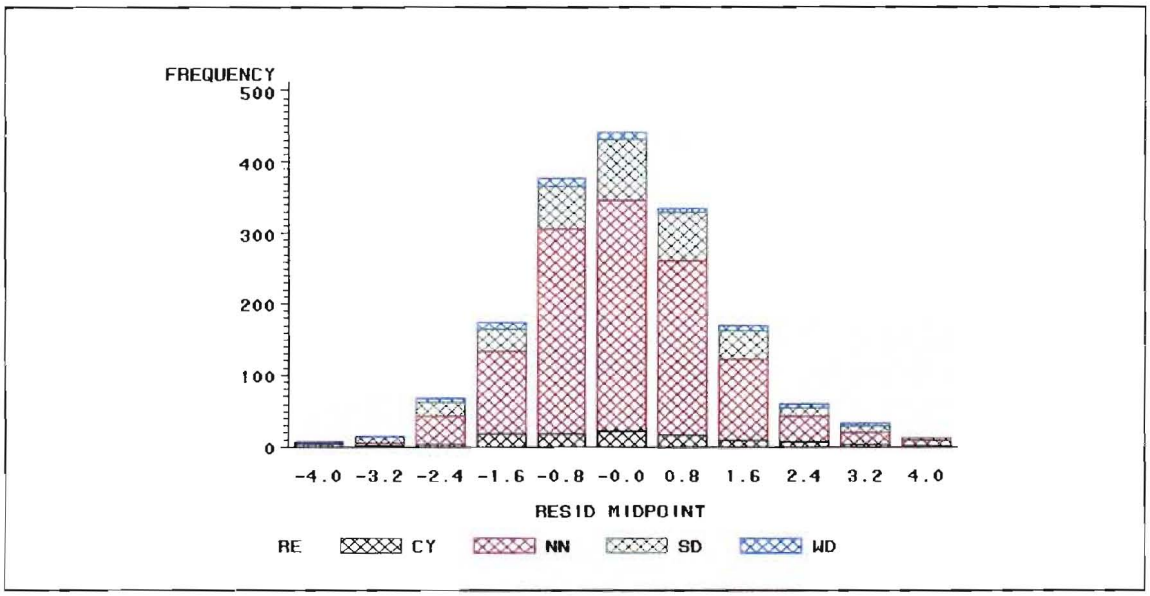

Figure 4.19: Frequency distribution of residuals for the long interval mean top height equation

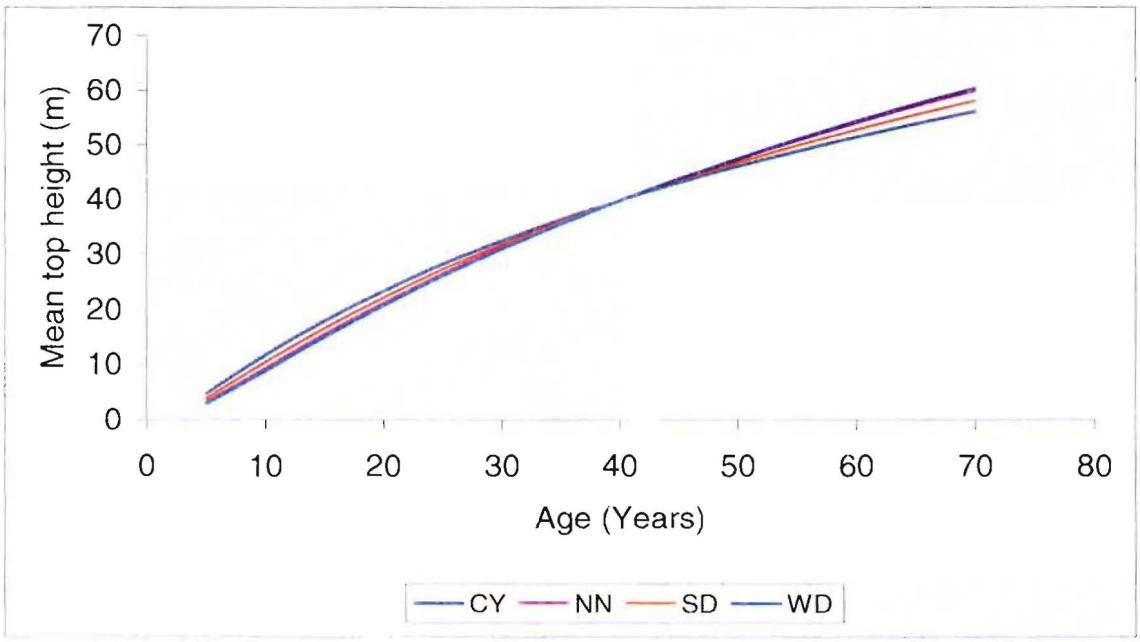

Figure 4.20: Height growth curves at site index = 40 for long interval data

4.2.4 Stem Survival/ha Equation

Mortality data used in the long interval data set were confined to those intervals where mortality occurred. Woollons (1998) pointed out that, if the data which exhibit no mortality over several years were discarded, then mortality should be overpredicted. However if observations with no mortality between re-measurements are retained, difficulties can occur in the residual plot analysis and fitting of the models. In order to overcome these problems, Woollons suggested the use of a two-stage regression process: firstly, mortality on a permanent sample plot over a given period of time can be regarded as a binomial outcome, and modelled by logistic regression (Hosmer and Lemeshow, 1989); secondly, one should model mortality using only the data which exhibits some mortality through a projection model to estimate stem survival/ha at time T.

In this study, however, the above method could not be applied, because the data should be prepared with a constant interval, otherwise bias in prediction may occur. In this particular data set, the intervals ranged from 8 to 28 years to predict long term effects. Therefore in this study the method adopted was to include intervals only where mortality occurred, as previous researchers have done (e.g.Temu, 1992; Ngugi 1996). The reason for this was to recognise that the response variable theoretically takes only discrete values and that by removing non-mortality measurement pairs a smoother more continuous nature is obtained, making the fitting of a model easier.

The data used to build the mortality equation are summarized in Table 4.18. The best model in the long interval data set was in equation 4.15 which is a modified Gompertz function.

$$N_2 = N_1 e(\beta(T_2^2-T_1^2))\, e(\alpha+\beta_1K_1+\beta_2K_2+\beta_3K_3+\beta_4S)(1-\exp(\beta(T_2^2-T_1^2))) \quad (4.15)$$

Table 4.18 Summary data for long interval stem survival/ha equation

Region	N	Variables	Mean	Minimum	Maximum
Canterbury	94	Age(yrs)	26.7	15.0	45.2
		Stems/ha	1464	450	2970
		Site index (m)	27.4	23.5	31.2
Nelson	643	Age(yrs)	27.6	7.0	59.0
		Stem/ha	1581	161	3533
		Site index(m)	33.2	25.1	37.3
Southland	272	Age(yrs)	31.0	7.7	68.0
		Stem/ha	1105	148	3366
		Site index (m)	30.2	22.2	40.6
Westland	69	Age(yrs)	36.9	8.0	59.1
		Stem/ha	961	222	1827
		Site index (m)	30.1	20.6	37.4

Site index was found to infer different asymptotes representing different mortality trends among regions, and so three dummy variables were needed.

Table 4.19 shows the parameters for the equation for stem survival/ha, while graphs of the residuals, which lie mainly within ±218 trees per ha are shown in Figure 4.21. Some biases were present at higher stockings in the Nelson region, but confirmation that they might have come from coding faults could not be obtained because there were no original records. The plots of residuals against age and site index, and the frequency distribution chart are given in Figures 4.22, 4.23 and 4.24 respectively. The scatter of residuals was quite large because of the long projection intervals.

Table 4.19 Coefficients for the long interval stem survival/ha equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	7.410424692	0.23148155804	15557255	1078	14512.4
β	-0.000188194	0.00001120077			
β1	0.773054363	0.11048968612			
β2	0.588559838	0.11105171950			
β3	0.657015483	0.13674906685			
β4	-0.073839886	0.01016286966			

The survival/ha equation using long term projection interval data was found to be difficult to fit successfully. Equation 4.15 had a mean residual of 5 stems/ha. This means that overall the equation will underpredict, by 5 stems/ha. Equation 4.15 had a value of –0.01 for skewness and 0.73 for kurtosis. The equation gave a maximum residual of 390 stems/ha, a minimum residual of –431 stems/ha, while 95% of residuals lay within ±212 stems/ha.

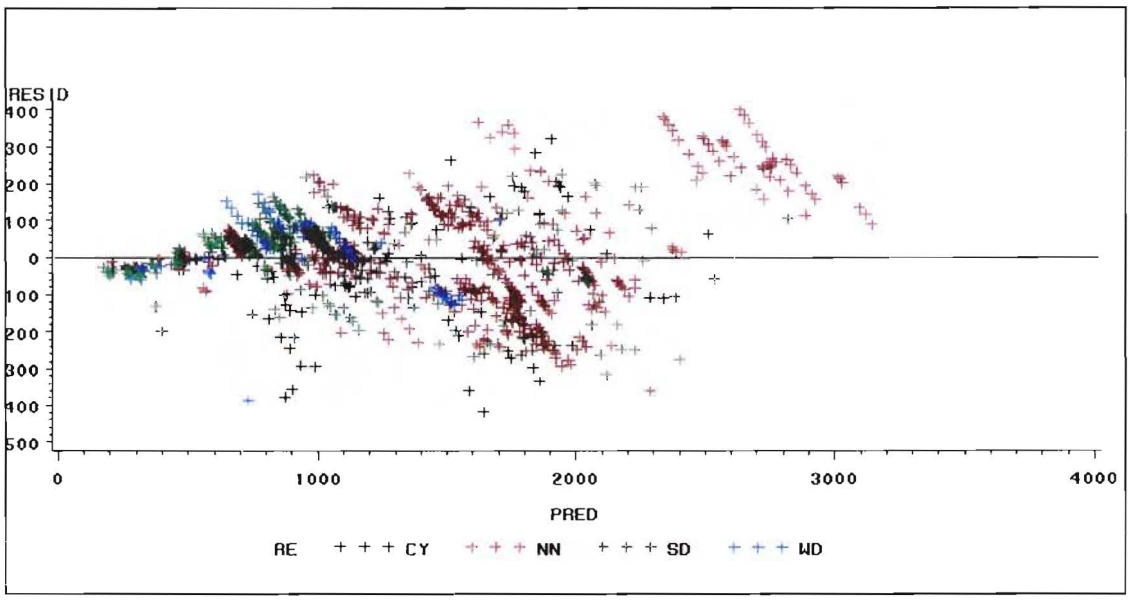

Figure 4.21: Plot of residuals vs predicted values for long interval stem survival/ha equation

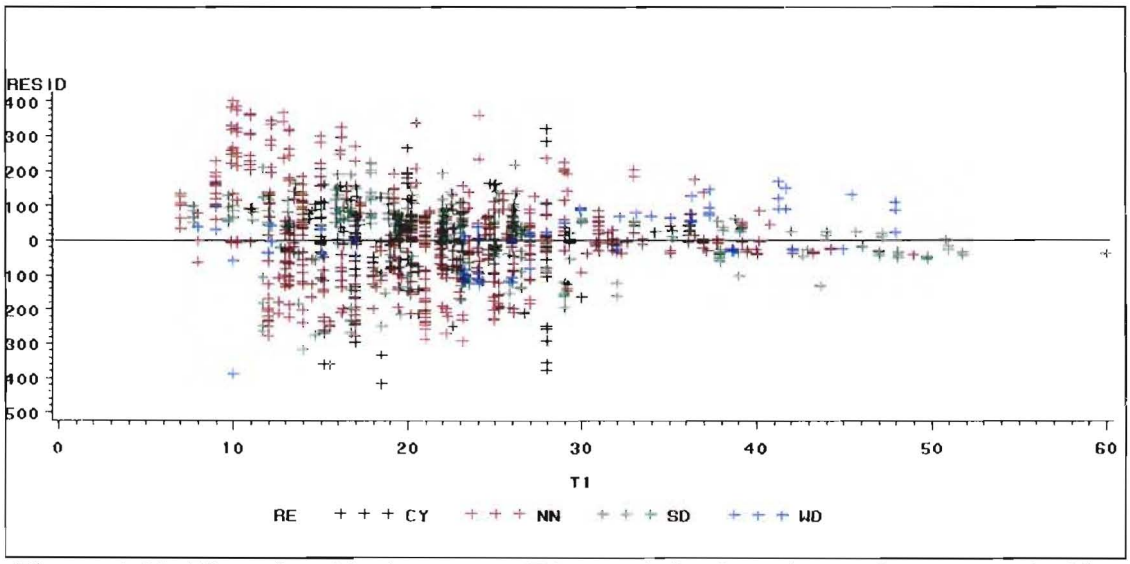

Figure 4.22: Plot of residuals vs age (T1 years) for long interval stem survival/ha equation

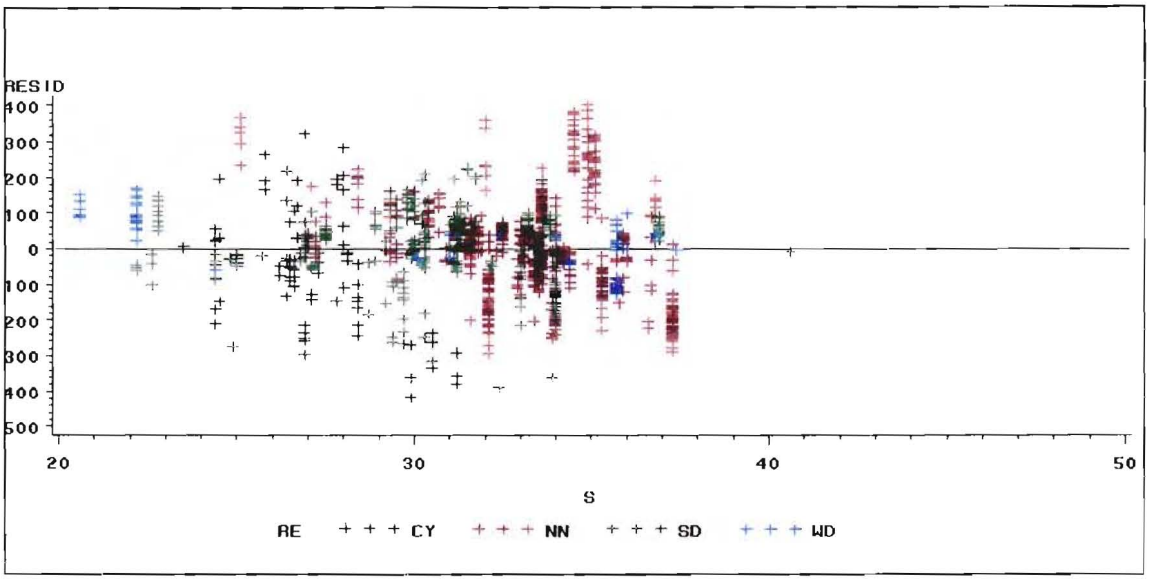

Figure 4.23: Plot of residuals vs site index for long interval survival stem/ha equation

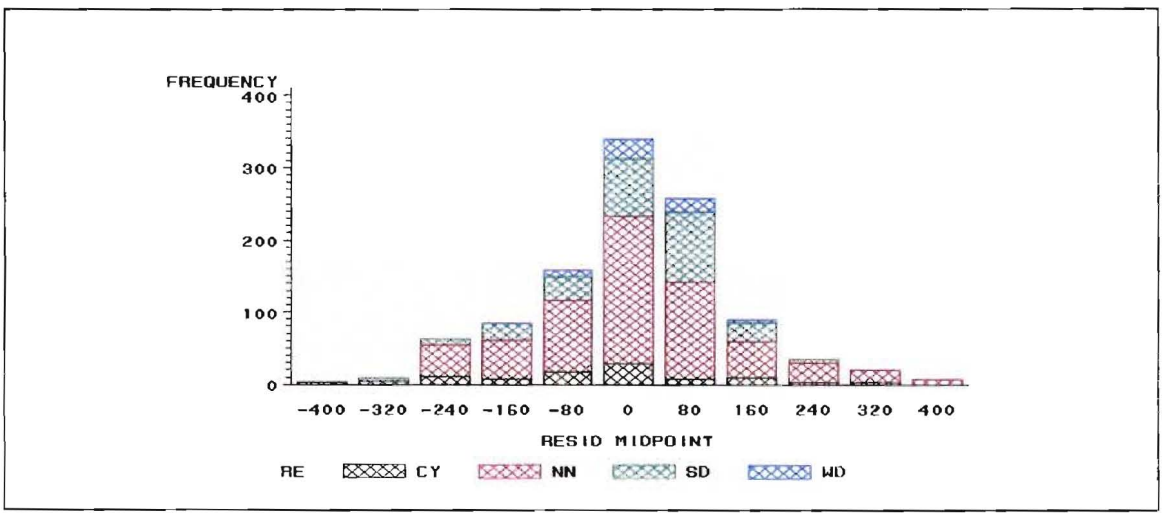

Figure 4.24: Frequency distribution of residuals for long interval stem survival/ha equation

4.3 MIXED INTERVAL DATA

4.3.1 Net Basal Area Projection Equation

The mixed interval data set consisted of merging the short and long interval data sets, which was done by PROC MERGE procedure in SAS (SAS Inc,1990). There were, therefore, more observations than for the short and long interval data sets individually. Its characteristics are summarized in Table 4.20.

Table 4.20 A summary of mixed intervals for basal area data

Region	N	Variables	Age (yrs)	G (m^2/ha)	Stems /ha	Site Index (m)	Alt (m)
Canterbury	165	Mean	26.9	40.5	1145	27.26	264
		Minimum	11.8	4.4	225	23.00	150
		Maximum	46.0	83.4	2970	31.20	470
Nelson	2505	Mean	26.0	40.4	962	32.96	465
		Minimum	7.0	1.2	93	24.60	130
		Maximum	59.1	115.3	3533	38.70	635
Southland	648	Mean	31.5	48.7	909	29.73	230
		Minimum	7.0	1.1	148	20.40	50
		Maximum	78.0	105.8	3366	40.60	625
Westland	373	Mean	29.5	32.6	737	31.69	251
		Minimum	5.0	0.01	198	20.60	0
		Maximum	59.1	121.0	4025	37.40	330

The same growth functions which were applied to the short and long interval data sets were fitted to the mixed interval set as done previously. The coefficients and the residual mean squares for candidate equations are presented in Table 4. 21.

Table 4. 21 Coefficients for general equations fitted to the mixed interval basal area /ha data

Model Name	Coefficients/Std.			MSE
	α	β	γ	
Hossfeld Polymorphic	99.6626/ 0.4976	-	2.8603/ 0.0158	27.375
Hossfeld Anamorphic	-	1.5627/ 0.1883	1.6338/ 0.0493	265.246
Schumacher Polymorphic I	5.1599/ 0.0049	-	-	28.034
Schumacher Polymorphic II	4.9584/ 0.0108	1.1992/ 0.0118	-	26.536
Schumacher Anamorphic	-	17.5404/ 0.7929	0.8016/ 0.0251	118.499
Chapman-Richards Polymorphic	0.8429/ 0.0411	0.5389/ 0.0305	0.0894 / 0.0098	31.928
Chapman-Richards Anamorphic	-	0.0542/ 0.0018	2.3592/ 0.0602	120.889
Gompertz Polymorphic	4.7809/ 0.0077	0.0916/ 0.0009	0.0006/ 0.0000	23.465
Weibull Polymorphic	1103.805/ 318.2224	0.0172/ 0.0044	0.5441/ 0.0192	30.424
Weibull Anamorphic	-	0.0032/ 0.0002	1.6943/ 0.0220	124.313

A Gompertz function, 4.16, gave the best residual pattern and also the lowest mean square error (MSE) for this data set, the same as for the long interval data set. The function was further examined by adding site variables, such as site index, altitude, thinning index and dummy variables.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{\alpha (1 - \exp (-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (4.16)$$

When site index was introduced to the basic Gompertz function, there was some improvement in the residual pattern, significant at the α =0.05 level, but the coefficient of site index was negative as found for the long interval data set.

Instead of site index, altitude was applied to equation 4.16, which gave greater reductions in the residual sums of squares (RSS) than site index. Altitude and thinning index were shown to be important explanatory variables for the best net basal area equation for short intervals together with dummy variables and all gave improvement to the fit when they were applied independently to equation 4.16. Therefore, modifications of the basic model through adding and subtracting these variables were applied to equation 4.16 to examine further possible improvement of the equation.

Addition of altitude and thinning terms gave some improvement but was not superior to the equation which included only the three dummy variables and altitude. The four regions Canterbury, Nelson, Southland and Westland were each seen to exhibit different growth patterns when represented with dummy variables K1, K2 and K3, as shown in equation 4.17.

The asymptote α in equation 4.17 is the default dummy variable representing the Canterbury region. When projections are made for Nelson, the asymptote is α $+\beta 1K1$, while those for Southland and Westland are given by $\alpha +\beta 2K2$ and $\alpha +\beta 3K3$ respectively.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (4.17)$$

Equation 4.17 appeared from the analysis to be the model which gave the best fit. Table 4.22 shows successive improvement from successive additional variables and the coefficients for this equation are presented in Table 4.23. The plots of the residuals against predicted values, age and altitude are shown in Figures 4.25, 4.26 and 4.27 respectively. The chart of residuals is shown in Figure 4.28.

Table 4.22 Successive improvement in fitting for mixed interval net basal area/ha equation

Input Variables	Error SS	% Reduction in ESS	MSE
Basic form (B)	87124.68	-	23.47
B, ALT	86999.65	0.14	23.44
B, Xt	86962.22	0.04	23.427
B, ALT, Xt	86867.43	0.11	23.408
B,Dummy(K1, K2, K3)	76963.51	11.4	20.75
B, Dummy and ALT	69004.11	10.3	18.73

Table 4.23 Coefficients for mixed interval net basal area/ha equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	4.470945799	0.01726794019	69004.11	3691	18.73
β	0.094444540	0.00081209371			
γ	0.000604193	0.00001123494			
$\beta 1$	0.180334470	0.01606062417			
$\beta 2$	0.342220416	0.01570185076			
$\beta 3$	0.342220146	0.02008447896			
$\beta 4$	0.000215222	0.00003056222			

The residual patterns were generally uniformly distributed without any serious bias. The mean average residual for equation 4.17 was –0.065 m^2/ha, indicating a slight overestimation. Skewness and kurtosis values were 0.031 and 2.46 respectively. This equation had a maximum residual of 15.73 m^2/ha, minimum residual of –16.25 m^2/ha, while 95% of residuals were contained within ±7.95 m^2/ha.

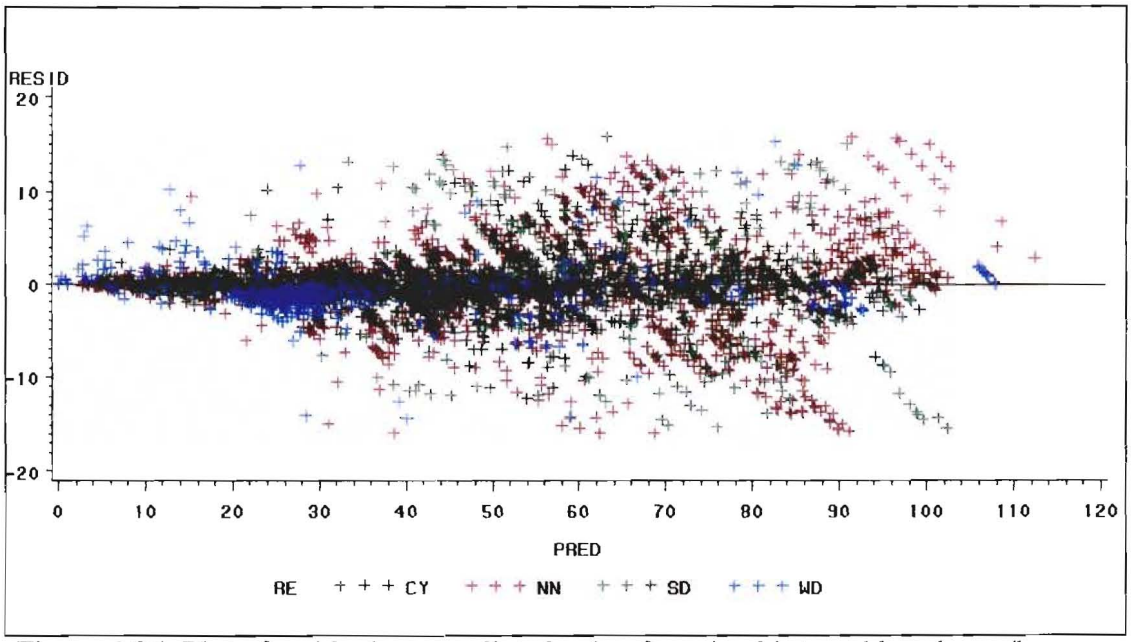

Figure 4.25: Plot of residuals vs predicted value for mixed interval basal area/ha equation

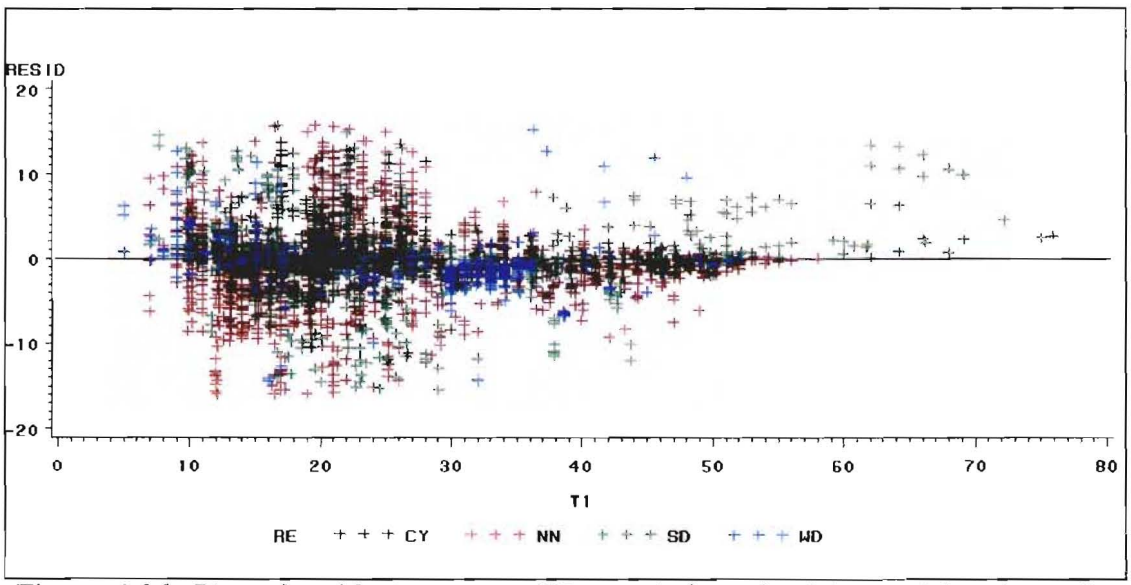

Figure 4.26: Plot of residuals vs age (T1 years) for mixed interval basal area/ha equation

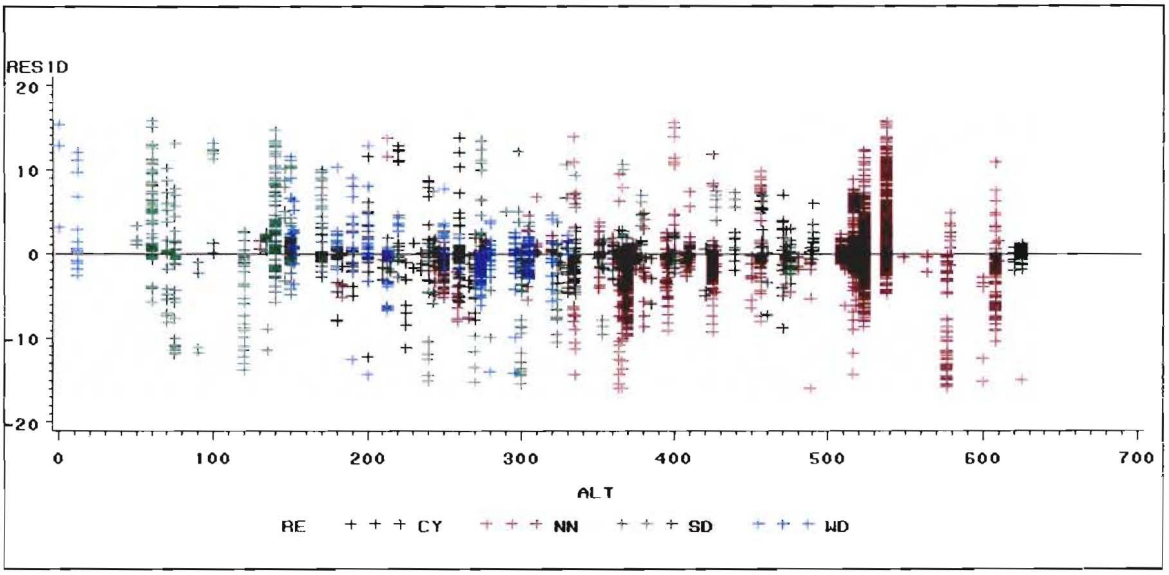

Figure 4.27: Plot of residuals vs altitude for the mixed interval data net basal area/ha equation

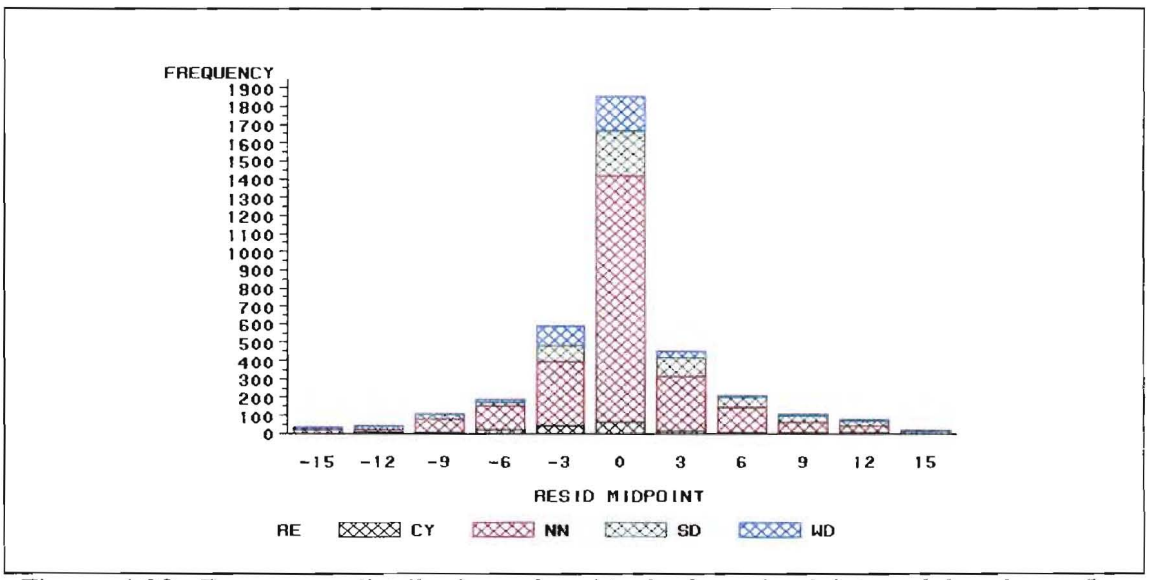

Figure 4.28: Frequency distribution of residuals for mixed interval basal area/ha equation

4.3.2 Mean Top Height Equation

The same procedures which were used in modelling mean top height in the short and long interval data sets were adopted for modelling with this data set. The mixed interval data used for mean top height are summarized in Table 4.24.

Table 4.24 Summary of data for mixed interval mean top height equation

Region	No.	Variable	Mean	Minimum	Maximum
Canterbury	160	Age(yrs)	27.0	11.8	45.2
		h_{100} (m)	19.5	8.3	34.6
Nelson	2491	Age(yrs)	25.8	7.0	59.1
		h_{100}(m)	21.8	5.6	45.0
Southland	642	Age(yrs)	31.3	7.0	78.0
		h_{100} (m)	22.6	4.3	45.8
Westland	379	Age(yrs)	29.9	5.0	59.1
		h_{100} (m)	22.3	1.9	42.1

A Schumacher polymorphic form was selected as the best equation and used for further examination after preliminary analysis with a variety of growth functions. Table 4.25 shows the coefficients and mean residual error for applied functions.

Table 4. 25 Coefficients for equations fitted to the mixed interval mean top height data

Model Name	Coefficients/Std.			MSE
	α	β	γ	
Hossfeld Polymorphic	79.6702/ 1.1231	-	1.4224/ 0.0074	1.013
Hossfeld Anamorphic	-	2.0057/ 0.0401	1.2371/ 0.0090	3.907
Schumacher Polymorphic I	4.0544 0.0034	-	-	2.374
Schumacher Polymorphic II	5.8182/ 0.0543	0.3701/ 0.0071	-	1.005
Schumacher Anamorphic	-	8.9346/ 0.0208	0.2993/ 0.0062	1.173
Chapman-Richards Polymorphic	0.7297/ 0.0500	0.2693/ 0.0394	0.0364/ 0.0040	1.380
Chapman-Richards Anamorphic	-	0.0220/ 0.0004	1.3875/ 0.0081	1.171
Gompertz Polymorphic	4.1796/ 0.0228	0.0544/ 0.0007	0.0003 / 0.0000	1.266
Weibull Anamorphic	-	0.0064/ 0.0000	1.2952/ 0.0050	1.184

Further analysis, adding altitude and dummy variables, was tried. Dummy variables showed that there were differences in height growth for all four regions. Altitude was also found to improve the fit on its own but not in conjunction with the dummy variables, and so was not included in the final equation selected. Therefore, the best fitting model took the form of equation 4.18.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/T_2)^{\beta}) \quad (4.18)$$

Parameter estimations for equation 4.18 are given in Table 4.26. The graphical plots of actual residuals against predicted values and age are shown in Figures 4.29 and 4.30. The chart of residuals is shown in Figure 4.31.

Table 4.26 Coefficient for mean top height equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	5.503163305	0.05267732000	3084.4165	3506	0.881
β	0.370593157	0.00672826534			
$\beta 1$	0.351539485	0.02537603196			
$\beta 2$	0.184418598	0.027386172			
$\beta 3$	0.399512718	0.03603392403			

Equation 4.18 had a mean residual error of 0.007 m and corresponding skewness and kurtosis values of 0.308 and 1.229 respectively. The equation gave a maximum residual of 3.3 m, a minimum residual of –3.4 m, while 95% of residuals lay within ±1.67 m.

4.3.3 Site Index Equation

The mean top height equation of form 4.18 was used to derive a site index equation by setting $h_{100,2}$ equal to site index (S) when T_2 equals 40 years.

$$S=h_{100,1}(T_1/40)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/40)^{\beta}) \quad (4.19)$$

Equation 4.19 could be rearranged to provide site index curves by making $h_{100,1}$ the subject. A series of site index curves for each region could be developed separately from the same equation by substituting the site index (S) and T_2 with the required values. Figure 4.32 represents an illustration of a series of site index curves for the mixed interval data at the index age of 40 years.

Height growth trends of Douglas-fir growing in the Nelson and Westland regions were found to have higher asymptotes than for the other two regions, while the Canterbury region was found to attain a lower asymptote than for all the other regions.

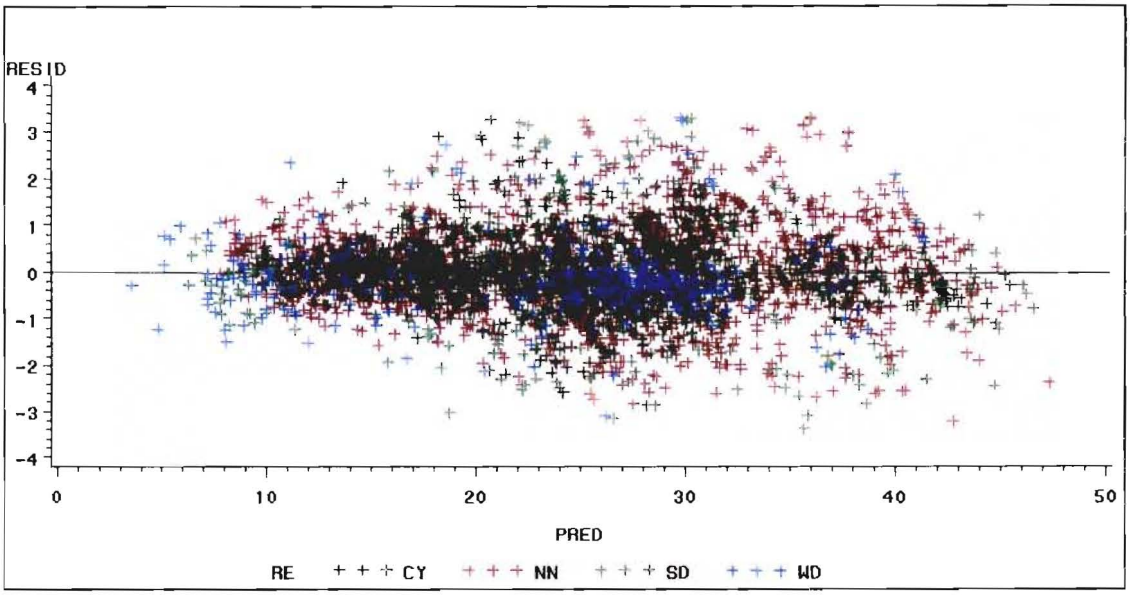

Figure 4.29: Plot of residuals vs predicted value for mixed interval mean top height equation

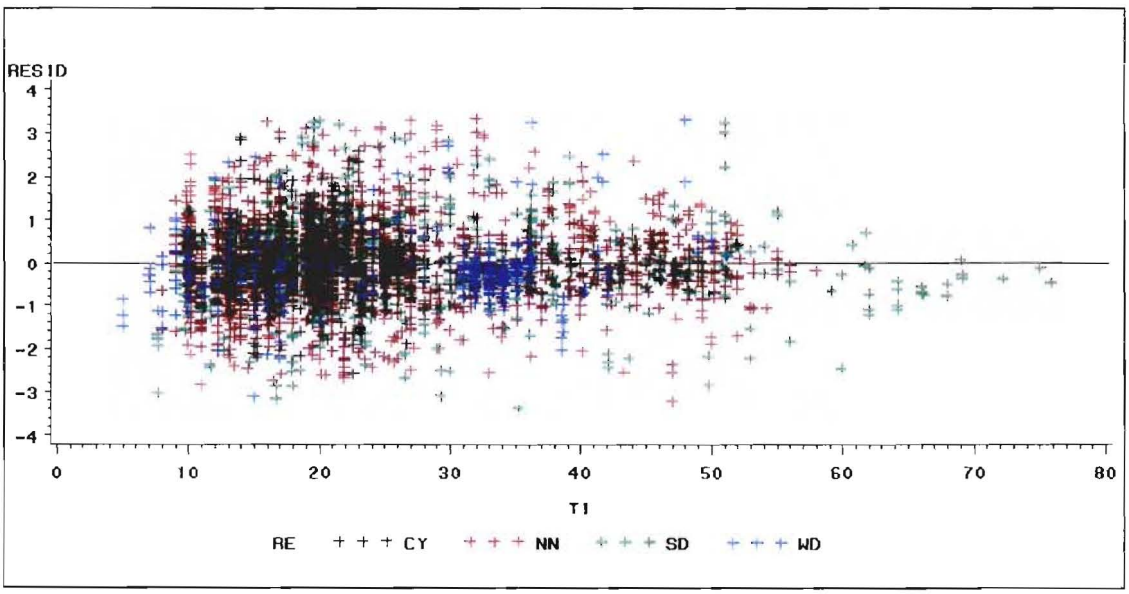

Figure 4.30: Plot of residuals vs predicted age (T1 years) for mixed interval mean top height equation

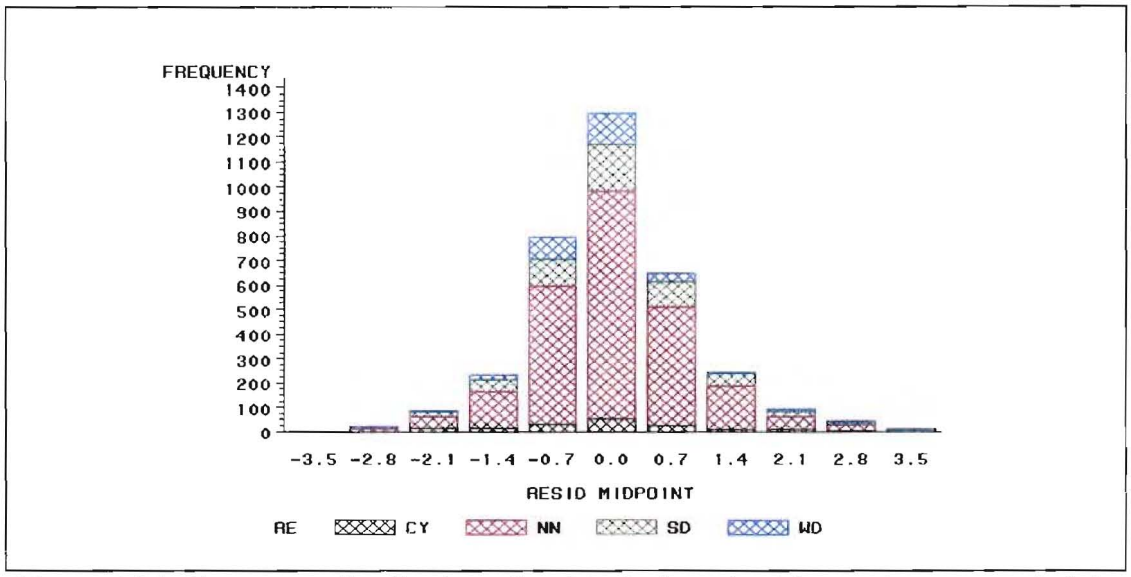

Figure 4.31: Frequency distribution of residuals for mixed interval mean top height equation

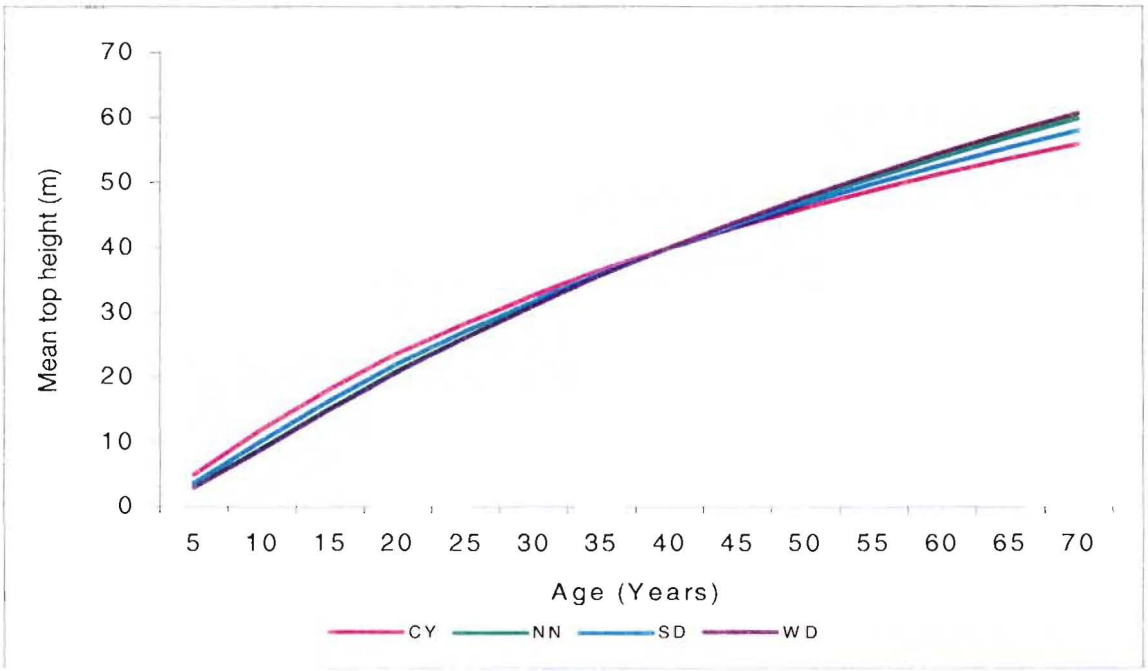

Figure 4.32: Height growth curves for site index = 40 for Douglas-fir

4.3.4 Stem Survival/ha Equation

The data used to build a mortality equation are summarized in Table 4.27. The best stems/ha model in the mixed interval data set was a modified Gompertz function, 4.20, which is the same form of equation shown to be best for the long interval data set.

Table 4.27 Summary data for mixed interval stem survival/ha equation

Region	N	Variable	Mean	Minimum	Maximum
Canterbury	108	Age (yrs)	27.5	15.0	45.2
		Stems/ha	1416	225	2970
		Site index (m)	27.35	23.00	31.20
Nelson	950	Age(yrs)	27.6	7.0	59.0
		Stem/ha	1601	161	3533
		Site index (m)	33.30	25.10	37.30
Southland	327	Age(yrs)	30.4	7.7	68.0
		Stem/ha	1141	148	3366
		Site index (m)	30.15	20.40	40.60
Westland	148	Age(yrs)	28.7	5.0	59.1
		Stem/ha	1204	222	4025
		Site index (m)	31.58	20.60	37.40

When modelling stem survival/ha, the data were again screened to ensure that only measurement pairs which did not contain a change in stocking were removed. Because such measurement pairs abound, the result of modelling with them included may be biased towards zero mortality.

$$N_2 = N_1 e^{(\beta(T_2^2-T_1^2))} e^{(\alpha+\beta_1K_1+\beta_2K_2+\beta_3K_3+\beta_4S)(1-\exp(\beta(T_2^2-T_1^2)))} \quad (4.20)$$

Application of dummy variables to examine differences of mortality across the four regions showed that there were significant differences in mortality trends. Site index also improved the model, while the negative coefficient for site index indicated that

trees in stands with higher site indices tended to suffer more mortality. However, some bias was detected for the Southland region and at higher stems/ha in the Nelson region. Table 4.28 shows the coefficients for the stem survival/ha equation.

Figure 4.33 shows the plot of residuals against predicted values. The plots of the residuals against age, site index and the frequency distribution chart are given in Figures 4.34, 4.35 and 4.36 respectively.

Table 4.28 Coefficients for mixed interval stem survival/ha equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	7.428277280	0.19364696547	16101160.0	1533	10544.3
β	-0.000190980	0.00000951205			
$\beta 1$	0.744421473	0.09233006941			
$\beta 2$	0.558021915	0.09255670168			
$\beta 3$	0.615293360	0.11361061080			
$\beta 4$	-0.072977005	0.00846269856			

The equation 4.20 had a mean residual of 1 stem/ha. The equation 4.20 resulted in a value of 0.07 for skewness and 1.53 for kurtosis. The equation gave a maximum residual of 374 stems/ha, a minimum residual of –426 stems/ha, while 95% of residual lay within ±177 stems/ha.

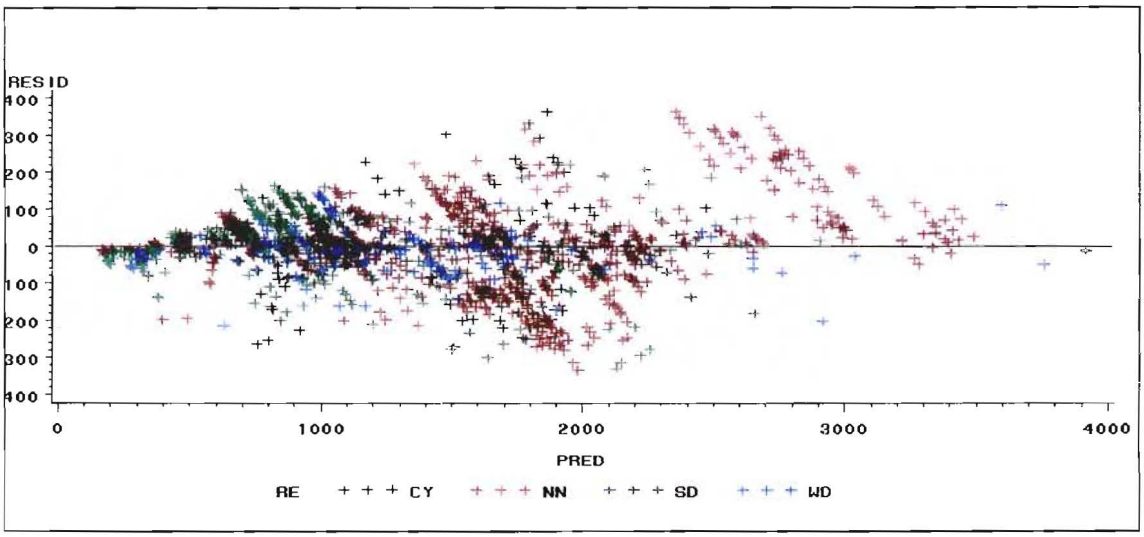

Figure 4.33: Plot of residuals vs predicted value for mixed interval stem survival /ha equation

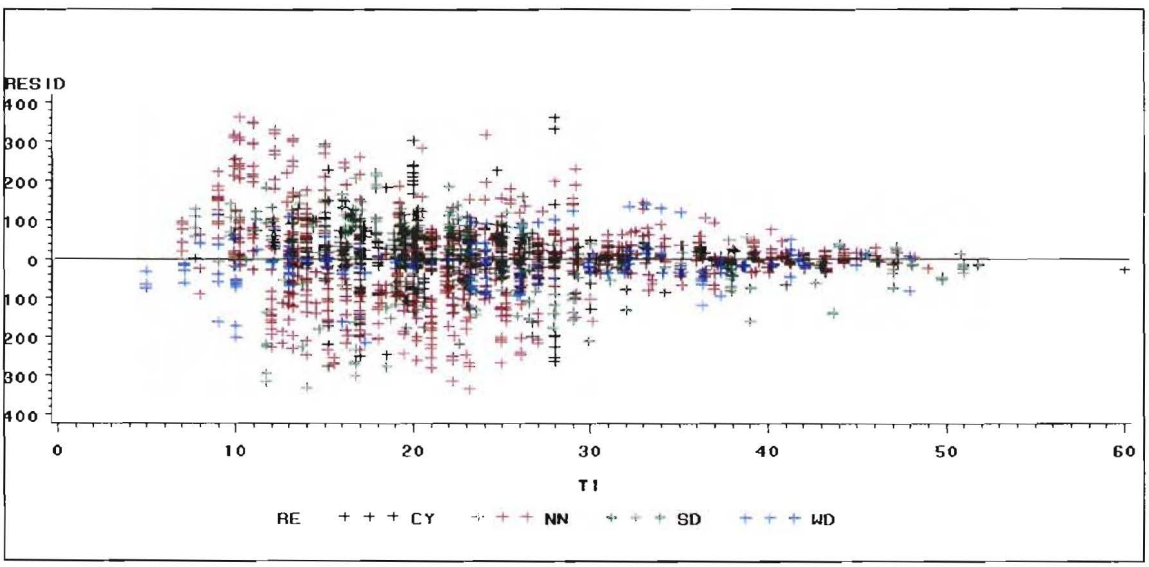

Figure 4.34: Plot of residuals vs age (T1 years) for mixed interval stem survival/ha equation

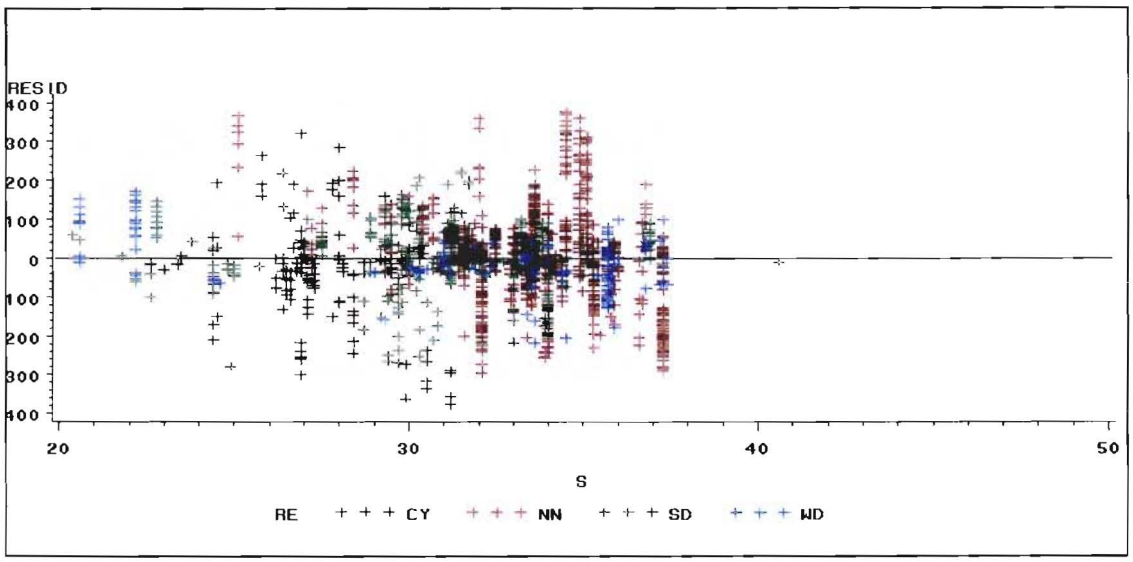

Figure 4.35: Plot of residuals vs site index for the mixed interval stem survival/ha equation

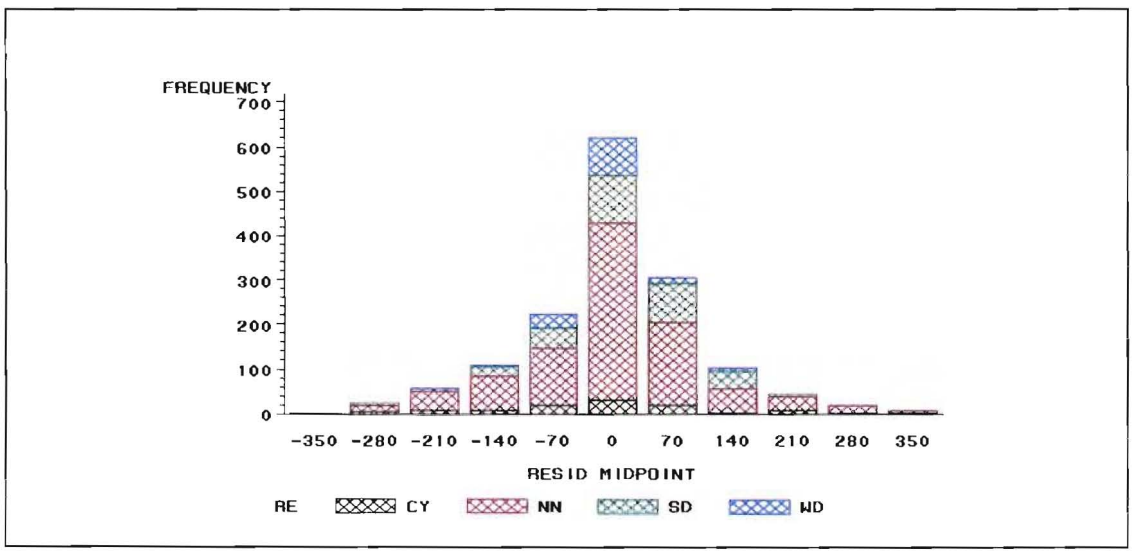

Figure 4.36: Frequency distribution of residuals for mixed interval stem survival/ha equation

4.4 Comparisons of Coefficients

In order to compare the precision of the coefficients between a model which was derived from the short interval data and a model which was obtained with the mixed interval data, a form of Schumacher model for the net basal area/ha and mean top height equations was tried for both data sets. The forms of equation applied to the data are presented in equations 4.21 and 4.22.

$$G_2=G_1(T1/T2)^{\beta}e(\alpha+ALT\beta_1+K_1\beta_2+K_2\beta_3+K_3\beta_4)(1-(T_1/T_2)^{\beta}) \quad (4.21)$$

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1)(1-(T_1/T_2)^{\beta}) \quad (4.22)$$

where G, h_{100} and ALT are basal area/ha, mean top height and altitude respectively, K1, K2 and K3 are dummy variables for location. After fitting the model coefficients were compared, as shown in Tables 4.29 and 4.30.

Table 4.29 Coefficients for net basal area/ha equation with the short and mixed interval data sets

Parameter	Estimates		Std.Error		95% Confidence Interval			
	Short data	Mixed data	Short data	Mixed data	Short data		Mixed data	
					Lower	Upper	Lower	Upper
α	4.5200	4.6848	0.0520	0.0225	4.418	4.622	4.641	4.729
β	1.1200	1.2207	0.0141	0.0110	1.092	1.148	1.199	1.242
β_1	0.0007	0.0001	0.0000	0.0000	0.001	0.001	0.000	0.000
β_2	0.1579	0.1559	0.0501	0.0203	0.060	0.256	0.116	0.196
β_3	0.5082	0.3514	0.0496	0.0199	0.411	0.605	0.312	0.390
$\beta 4$	0.4402	0.3799	0.0532	0.0258	0.336	0.544	0.329	0.430

Table 4.30 Coefficients for mean top height equation with the short and mixed interval data sets

Parameter	Estimates		Std.Error		95% Confidence Interval			
	Short data	Mixed data	Short data	Mixed data	Short data		Mixed data	
					Lower	Upper	Lower	Upper
α	5.2032	5.6877	0.0932	0.0508	5.020	5.386	5.588	5.787
β	0.4328	0.3682	0.0166	0.0068	0.400	0.465	0.355	0.382
β_1	0.1736	0.1851	0.0391	0.0120	0.097	0.250	0.162	0.208

The confidence intervals for both models showed that all estimates were significant at α=0.05 level. The 95 percent confidence interval for the mixed interval model for net basal area/ha and mean top height equations were much tighter. The t statistics were also much greater for the mixed interval data in both the net basal area/ha and mean top height equations. Though the t statistics could not be used explicitly because of the correlated data, they provide a measure of relative precision. General trends of the coefficient estimation in Tables 4.29 and 4.30 show that short interval models decrease the asymptote for both basal area/ha and mean top height equations, as well as increasing the slope coefficient for the mean top height equation and decreasing that for the basal area equation.

In order to test the long term projection capabilities of the two models, a long interval data set was used, in which intervals more than 8 years were contained. The trend of residual patterns of the mixed interval equation was tighter than that of the short interval equation for the long interval data. The extreme residuals for the mixed interval net basal area/ha equation were –22.85 and 21.67 m^2/ha, while corresponding values for the short interval equation were –26.67 and 26.42 m^2/ha. The mean residual value and skewness for the mixed interval equation were 0.62 m^2/ha and 0.11, which were better than those for the short interval equation with 0.69 m^2/ha and –0.21. The plots of residuals against predicted values are presented in Figures 4.37 and 4.38.

The frequency bar chart of residuals was well distributed for the mixed interval mean top height equation when applied to long interval data, but that for the short interval was imbalanced as shown in Figures 4.41 and 4.42. The mean residual value for the short interval mean top height equation was 0.44, about ten times greater than that for the mixed interval equation with 0.04. Ninety five percents of residuals were contained within ± 2.8 m for the short interval equation, while the corresponding value for the mixed interval equation was ± 2.3 m. The plots of residuals against predicted values are shown in Figures 4.39 and 4.40.

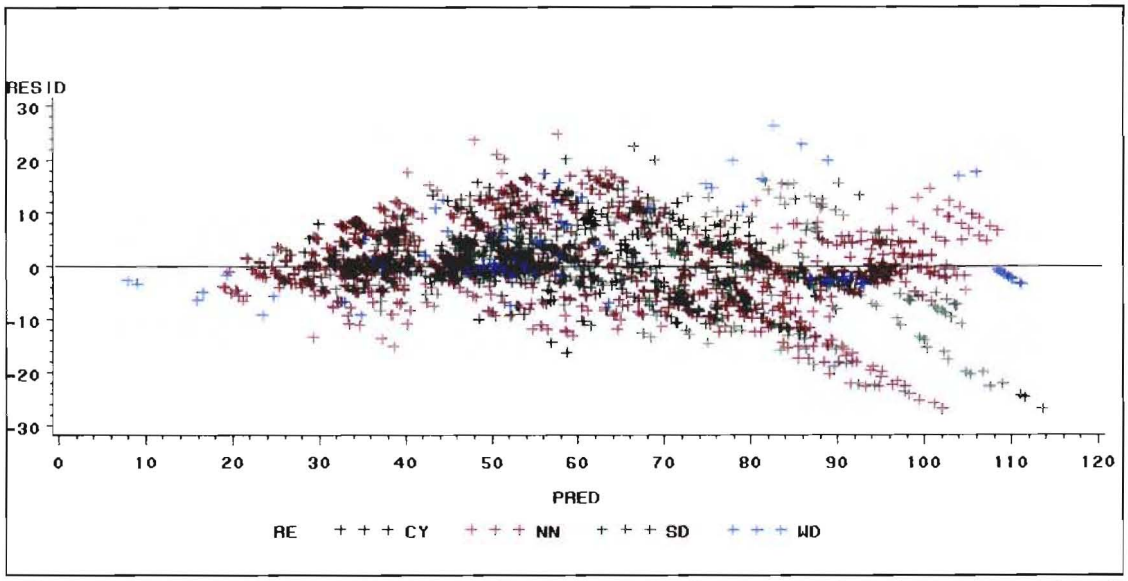

Figure 4.37: Plot of residuals vs predicted values in fitting the short interval net basal area/ha equation to long interval data to compare coefficients.

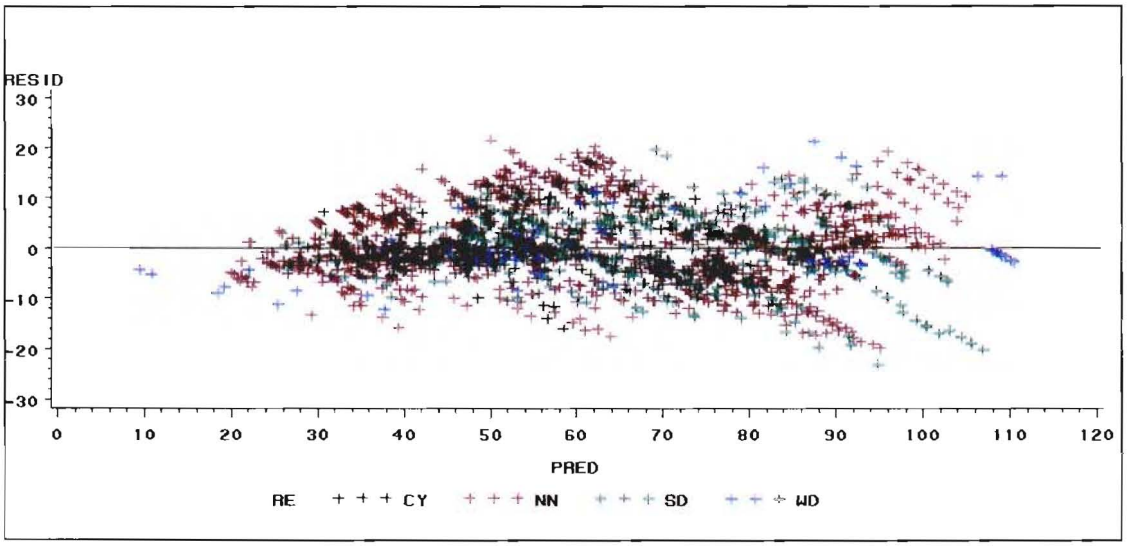

Figure 4.38: Plot of residuals vs predicted values in fitting the mixed interval net basal area/ha equation to long interval data to compare coefficients

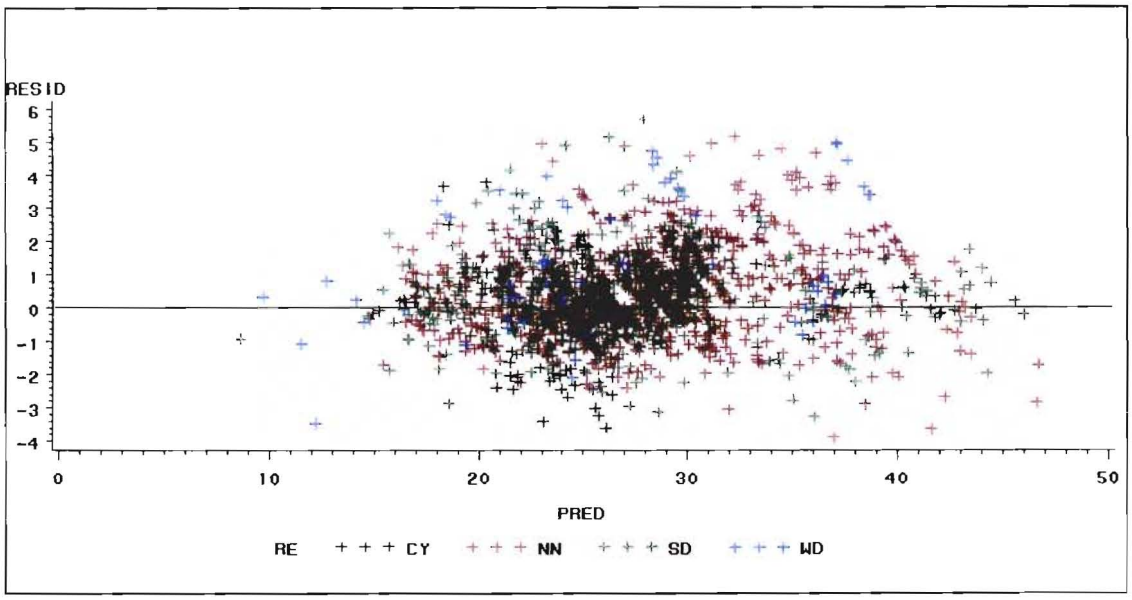

Figure 4.39: Plot of residuals vs predicted values in fitting the short interval mean top height equation to long interval data to compare coefficients

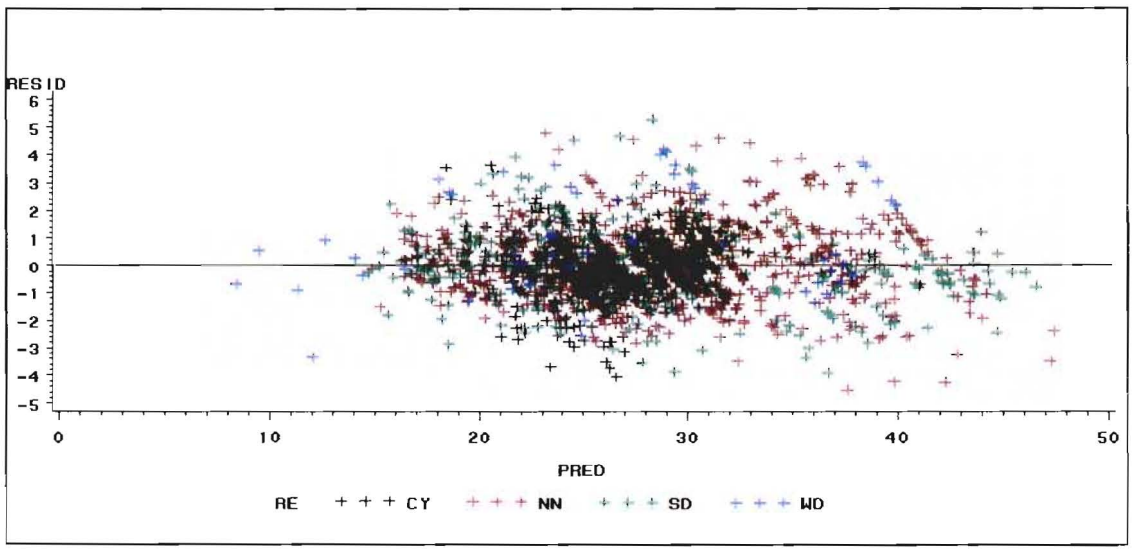

Figure 4.40: Plot of residuals vs predicted values in fitting the mixed interval mean top height equation to long interval data to compare coefficients.

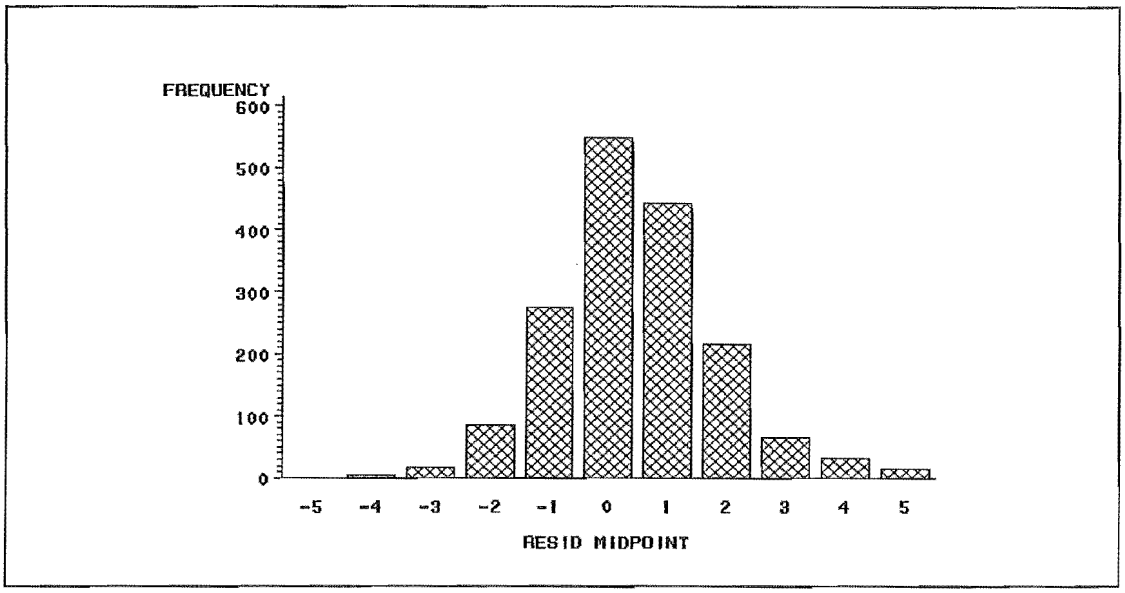

Figure 4.41: Frequency distribution of residuals in fitting the short interval mean top height equation to compare coefficients

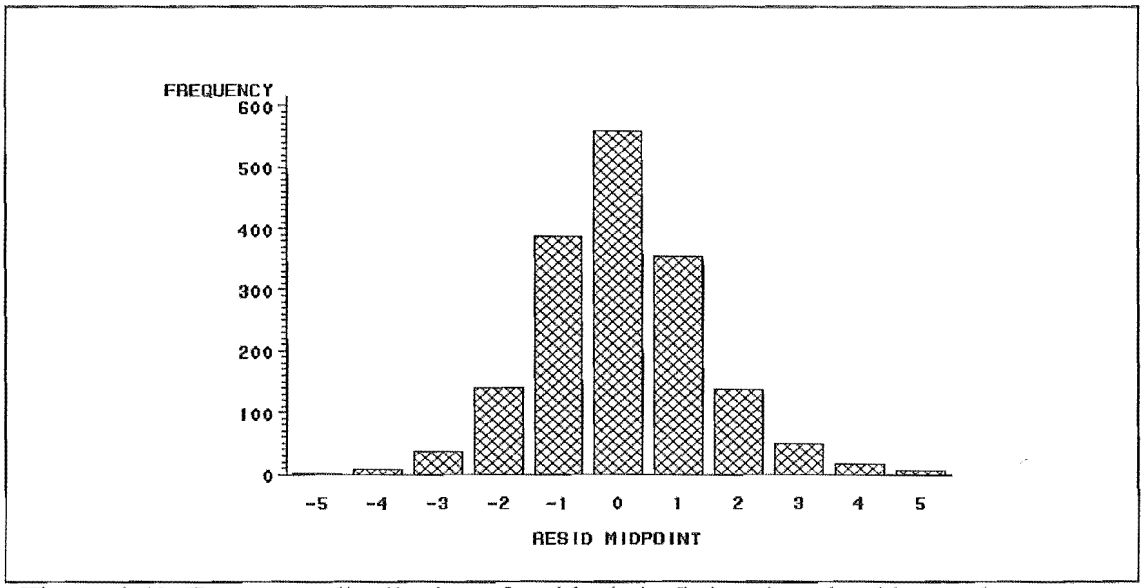

Figure 4.42: Frequency distribution of residuals in fitting the mixed interval mean top height equation to compare coefficients

Chapter 5

CROSS FITTING

In order to examine the extent of compatibility among the predicting equations developed for the three different interval data sets, each model was applied to each of the other data sets, and residual patterns were plotted to see the goodness of fit.

5.1 Net Basal Area Projection Equations

Six combinations of applying the coefficients from the three relevant equations which had been found best for each interval data set were evaluated. Accordingly, the best model in the short interval data set was applied to the long and mixed interval data sets; similarly, the best models in the long and mixed interval data were applied to the other two data sets.

5.1.1 Applying the best short interval equation to the other data sets

The best model for the short interval data was a Schumacher polymorphic function which included thinning index and three dummy variables as predictor variables as shown in equation 4.5.

$$G_2=G_1(T_1/T_2)^{\beta 1}\exp((\alpha+ALT\beta_2+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(T_1/T_2)^{\beta 1})+\beta_6 Xt(1/T_2-1/T_1)Tt/T_2) \quad (4.5)$$

The coefficients for this equation are in Table 4.5. When this function was applied to the long and the mixed interval data sets, the residual patterns were biased and showed less precision than the best equation specific to the long and mixed interval data sets.

The residuals showed that equation 4.5 overpredicted basal area increment at high values of basal area/ha. The distribution of residuals was positively skewed. Table 5.1 summarizes relevant statistics

Table 5.1 Summary of applying the best short interval equation to the other data sets

Variable	Long interval data set	Mixed interval data set
Mean residual error(m^2/ha)	0.66809	0.282207
95 % of residuals(m^2/ha)	±13.03	±9.89
Range of residuals(m^2/ha)	-29.5 / 26.6	-29.5 / 23.8

When this function was applied to the long and mixed interval data sets, the residuals lay within mainly ±13 m^2/ha and ±9.9 m^2/ha respectively. The plots of residuals against predicted values are shown in Figures 5.1 and 5.2.

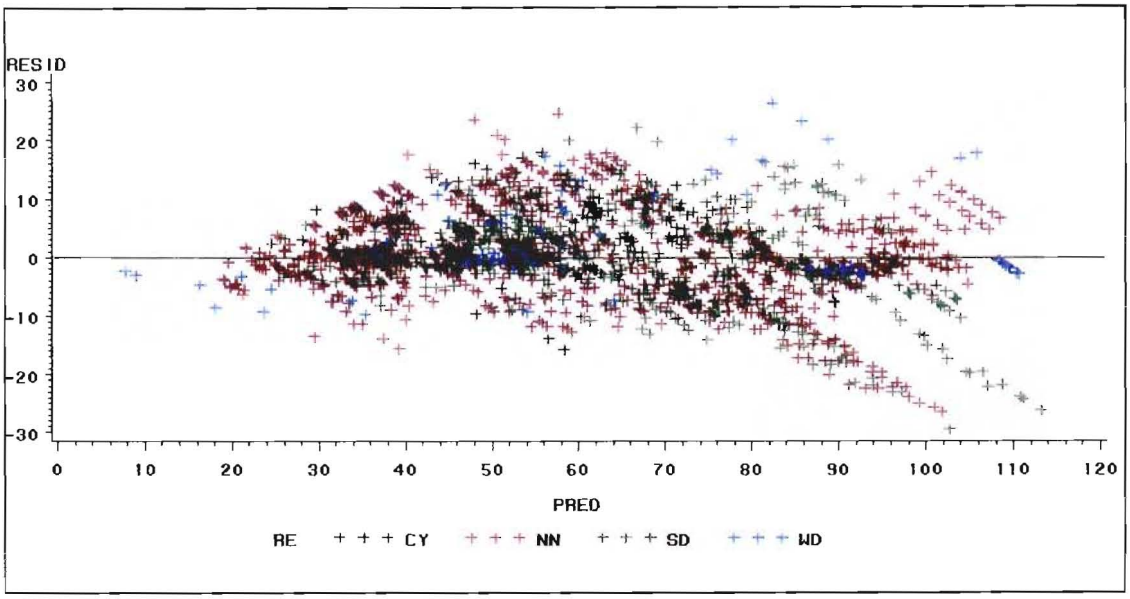

Figure 5.1: Plot of residuals vs predicted values when applying the best short interval equation for basal area/ha to the long interval data

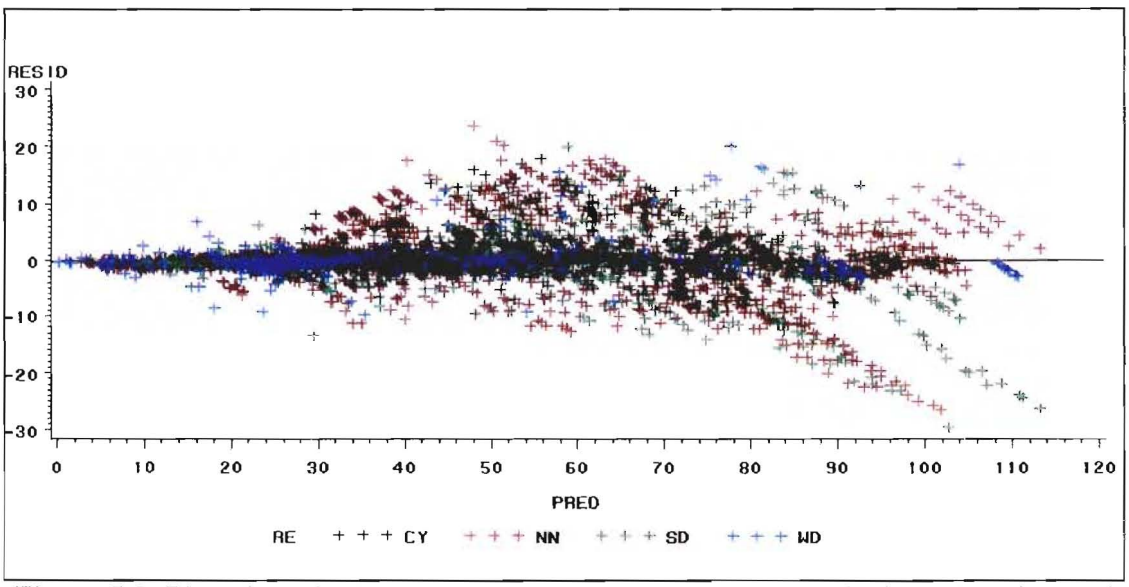

Figure 5.2: Plot of residuals vs predicted values when applying the best short interval equation for basal area/ha to the mixed interval data

5.1.2 Applying the best long interval equation to the other data sets

The Gompertz polymorphic function, equation 4.11 with altitude and three dummy variables for regions, was found to give the best fit for the long interval data set. The coefficients for this equation are in Table 4.14.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (4.11)$$

When the basal area/ha equation 4.11 was applied to the short interval data set, Westland region showed a slight imbalance of residuals at basal areas of 20 to 30 m^2/ha, but generally well balanced otherwise, though it was not as precise as the best model for the short interval data set. Applied to the mixed interval data set, the predictions were good without a biased residual pattern of residuals and showed as close correspondence to those of the best model for the mixed data set itself.

Mean residual error and skewness were -0.089 and 0.05 respectively, which were similar to the values for the best model for the mixed interval data set (-0.06 and 0.03). Table 5.2 shows a summary of relevant statistics.

Table 5.2 Summary of applying the best long interval equation to the other data sets

Variable	Short interval data set	Mixed interval data set
Mean residual error(m^2/ha)	-0.12039	-0.08932
95 % of residuals(m^2/ha)	±2.39	±8.17
Range of residuals(m^2/ha)	-6.19 to 6.79	-20.18 to 19.0

The plots of residual against predicted values are shown in Figures 5.3 and 5.4.

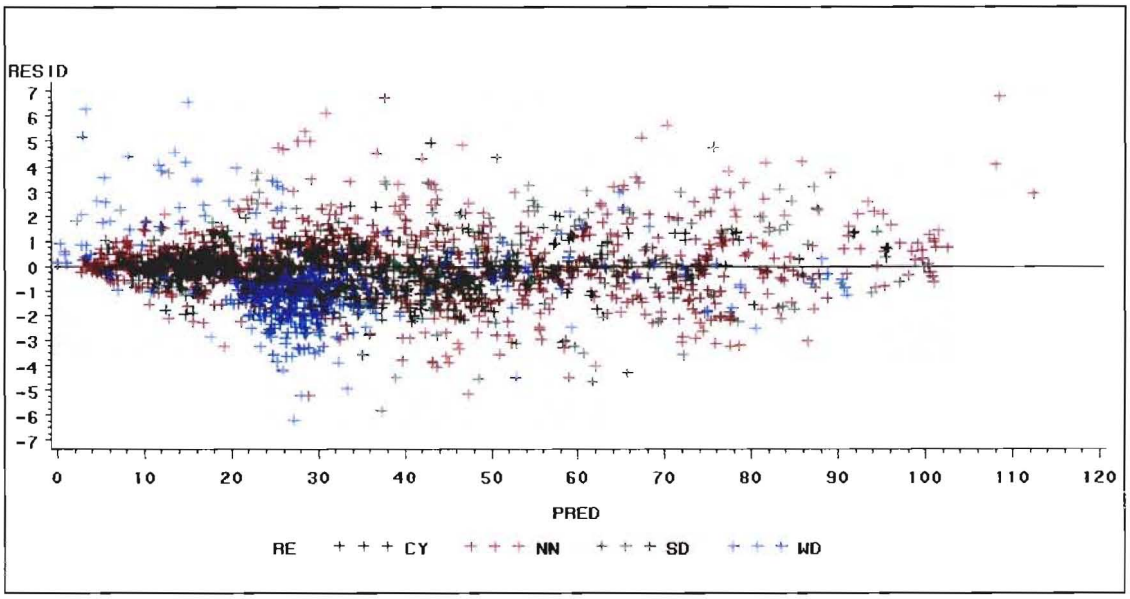

Figure 5.3: Plot of residuals vs predicted values when applying the best long interval equation for basal area/ha to the short interval data

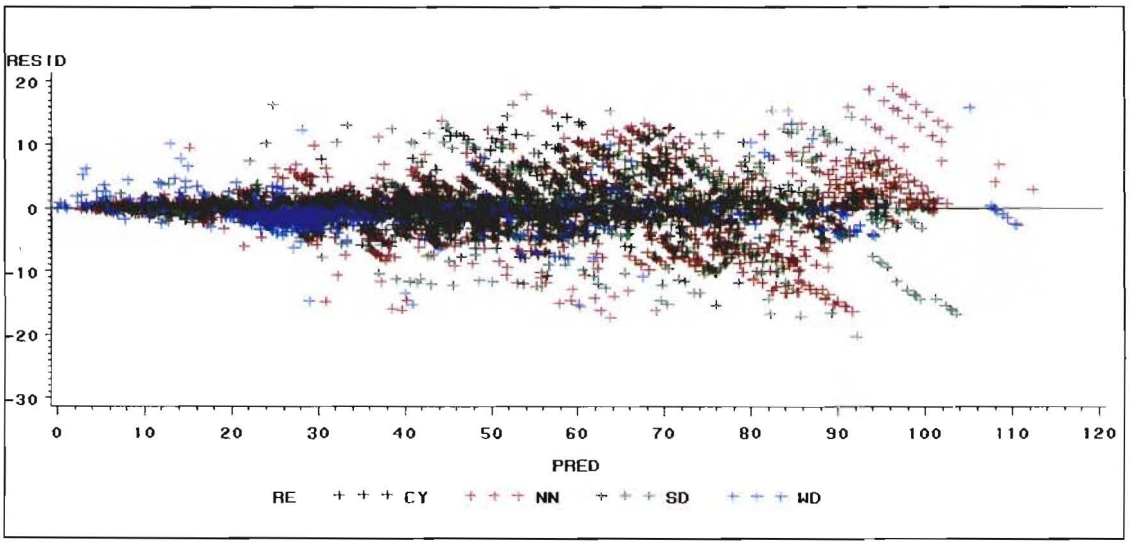

Figure 5.4: Plot of residuals vs predicted values when applying the best long interval equation for basal area/ha to the mixed interval data

5.1.3 Applying the best mixed interval equation to the other data sets

Basal area/ha prediction for the mixed interval data set was best represented by equation 4.17.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (4.17)$$

When equation 4.17 was applied to the short interval data set, the overall residuals were generally acceptable without apparent bias, though a slight bias for the Westland region was evident as shown in Figure 5.5. A similar pattern to that of applying the best long interval equation to the short interval data set was evident. Applied to the long interval data set, the predictions fitted well without a biased residual pattern, as shown in Figure 5.6. Table 5.3 presents a summary of relevant statistics.

Table 5.3 Summary of applying the best mixed interval equation to the other data sets

Variable	Short interval data set	Long interval data set
Mean residual error(m2/ha)	-0.10343	0.01845
95 % of residuals(m2/ha)	±2.43	±11.45
Range of residuals	-6.0 to 6.8	-21.7 to 22.9

When this function was applied to the short and long interval data sets, the residuals lay within mainly ±2.43 m^2/ha and ±11.45 m^2/ha respectively.

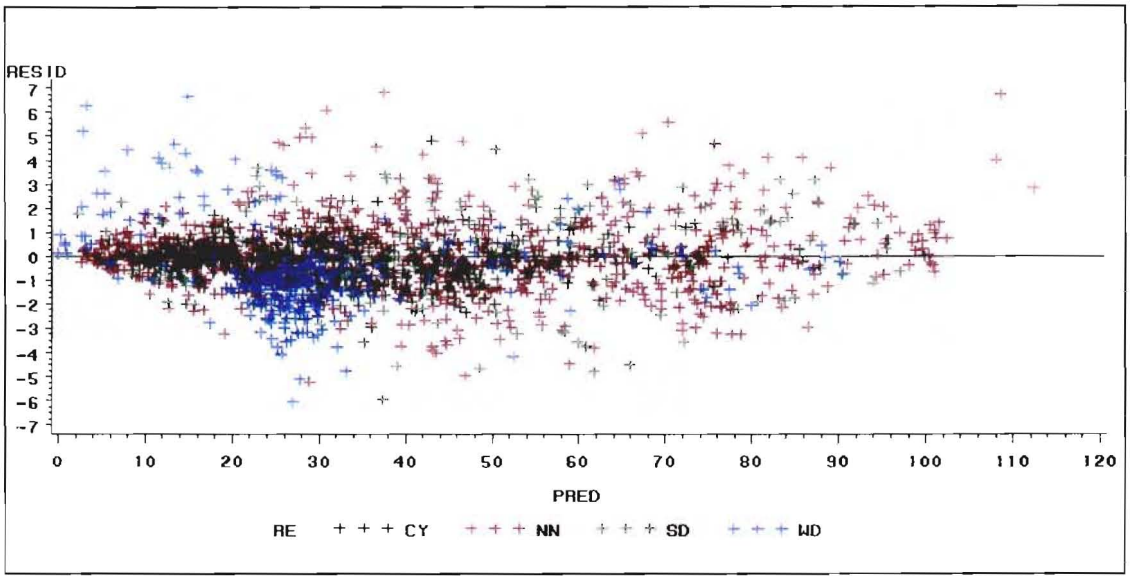

Figure 5.5: Plot of residuals vs predicted values when applying the best mixed interval equation for basal area/ha to the short interval data

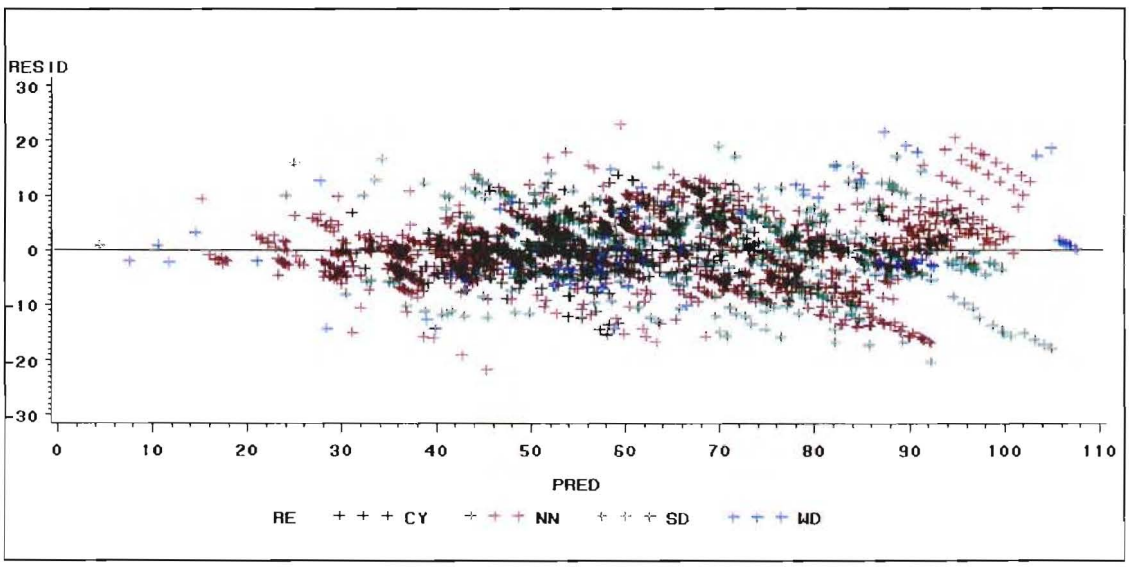

Figure 5.6: Plot of residuals vs predicted values when applying the best mixed interval equation for basal area/ha to the long interval data

5.2 Mean Top Height Equations

5.2.1 Applying the best short interval equation to the other data sets

The Schumacher polymorphic function 4.7 was found to give the best fit for the short interval data set.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+\beta_2 ALT+\beta_1 K_1)(1-(T_1/T_2)^{\beta}) \qquad (4.7)$$

The coefficients of this equation are in Table 4.8. When applied to the long and mixed interval data sets, the residual patterns were similar to those from applying the best functional forms for the long and mixed intervals. The plots of residual against predicted values are shown in Figures 5.7 and 5.8.

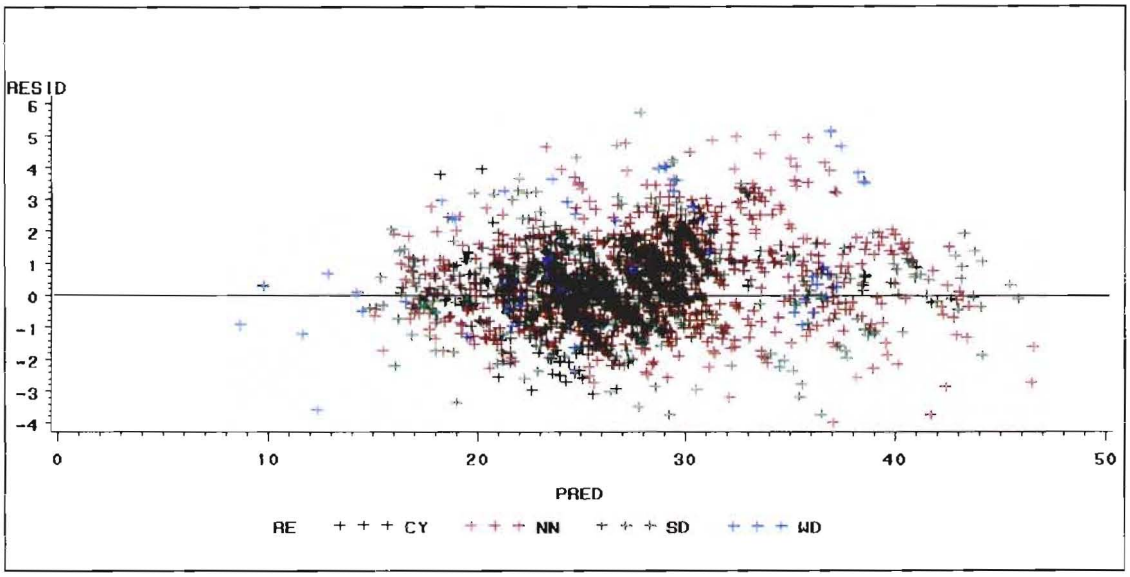

Figure 5.7: Plot of residuals vs predicted values when applying the best short interval equation for mean top height to the long interval data

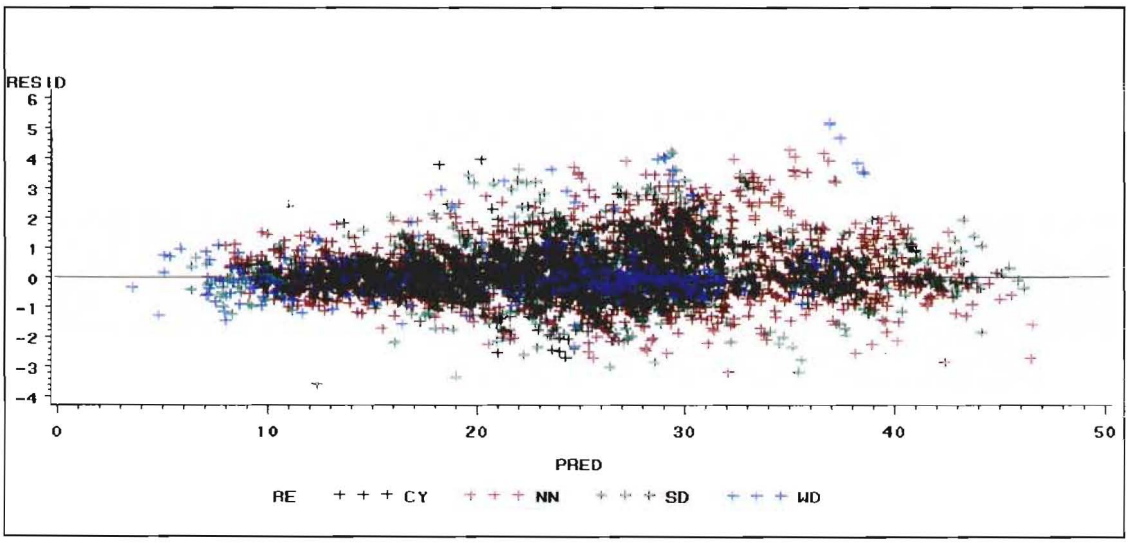

Figure 5.8: Plot of residuals vs predicted values when applying the best short interval equation for mean top height to the mixed interval data

5.2.2 Applying the best long and mixed interval equations to the short interval data set

The Schumacher polymorphic function 4.13 was found to give the best fit for both the long and the mixed interval data sets.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/T_2)^{\beta}) \quad (4.13)$$

The coefficients of this equation are in Tables 4.17 and 4.26. When equation 4.13 was applied to the short interval data set, the residuals were uniformly distributed without any serious bias, and the residual pattern was almost the same as that from applying equation 4.7.

When equation 4.13 was applied to the long and mixed interval data sets using their own coefficients, the residual pattern was nearly the same as for the best model for each interval data set, because the same functional form and similar coefficients were used for the two equations. These comparisons are summarised in Table 5.4.

Table 5.4 Summary of statistics applying the best mean top height equation for each data set to the other data sets

Variable	Short best	Long to short	Mixed to Short	Long best	Short to long	Mixed to long	Mixed best	Short to mixed	Long to mixed
Mean residuals	0.030	-0.030	-0.027	0.02	0.43	0.035	0.007	0.23	0.005
Skew ness	0.35	0.35	0.36	0.21	0.31	0.25	0.30	0.71	0.31
95% residuals	±1.0	±1.0	±1.0	±2.25	±2.89	±2.28	±1.67	±2.10	±1.68

The mean residual value for the short interval data set was 0.030, about one third greater than that for the mixed set of 0.007. The skewness for the short interval data had a value of 0.35 compared with 0.30. When the mixed interval model was applied to the short data set it produced almost the same values for mean residual error and

skewness as those for the short data best model, but the short model was not as good a fit for the mixed data set as shown in Table 5.4. The plots of residual against predicted values are shown in Figures 5.9, 5.10, 5.11 and 5.12.

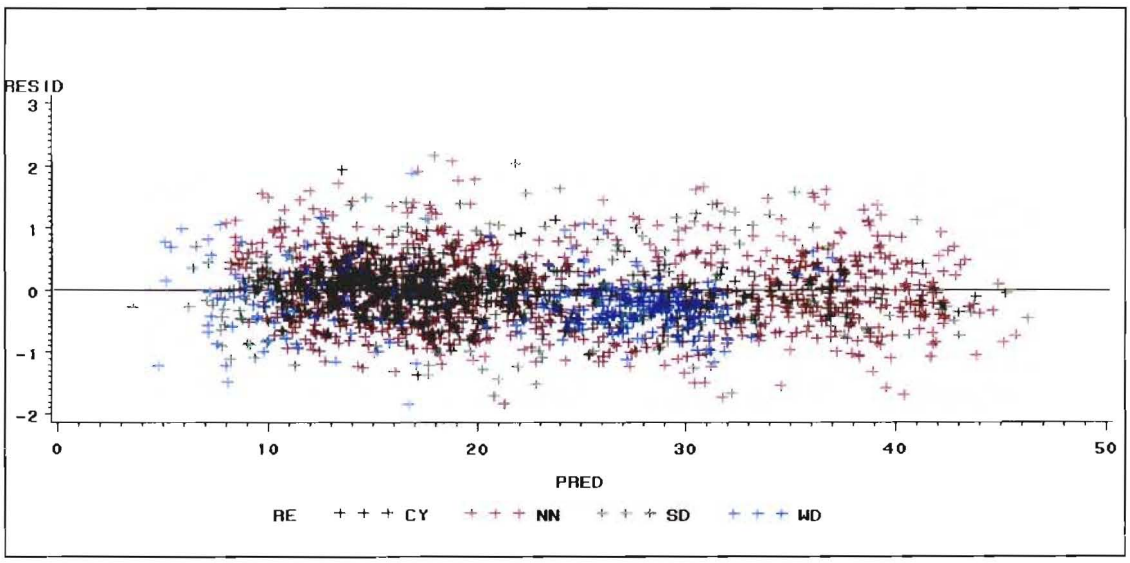

Figure 5.9: Plot of residuals vs predicted values when applying the best long interval equation for mean top height to the short interval data

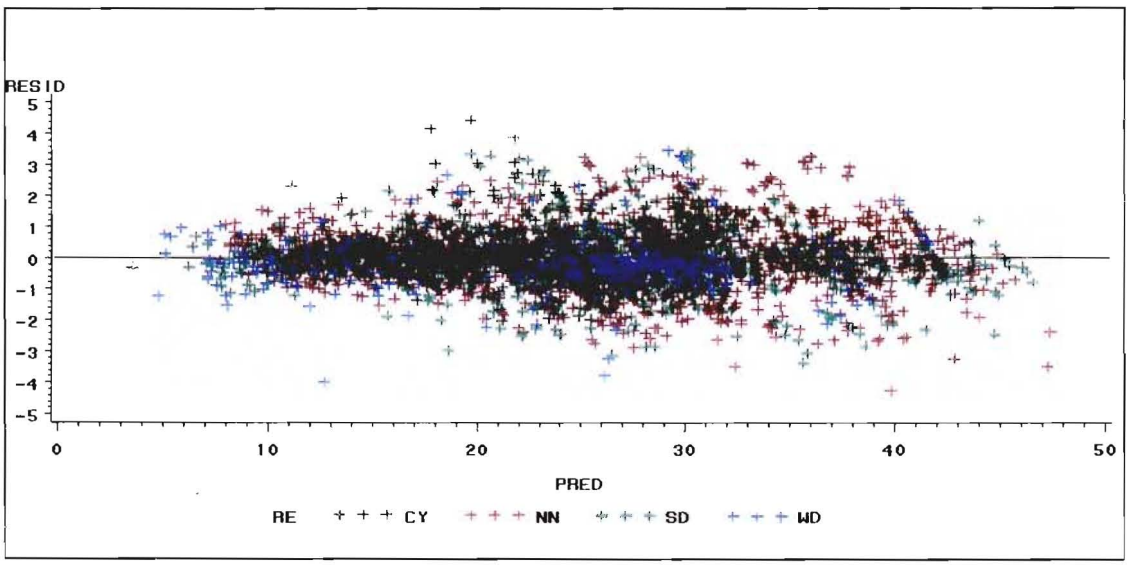

Figure 5.10: Plot of residuals vs predicted values when applying the best long interval equation for mean top height to the mixed interval data

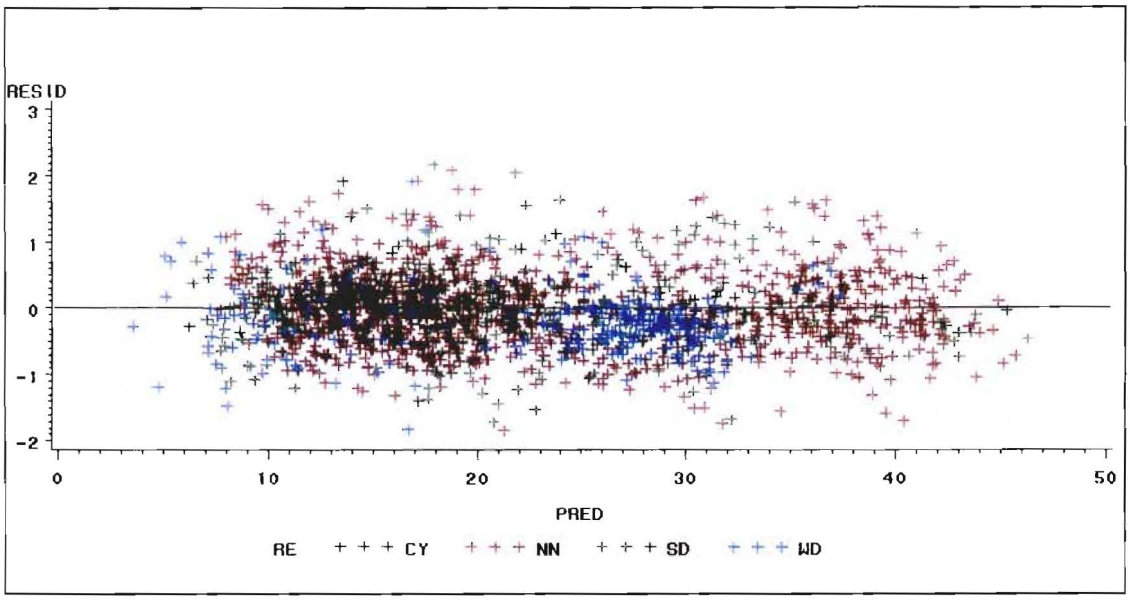

Figure 5.11: Plot of residuals vs predicted values when applying the best mixed interval equation for mean top height to the short interval data

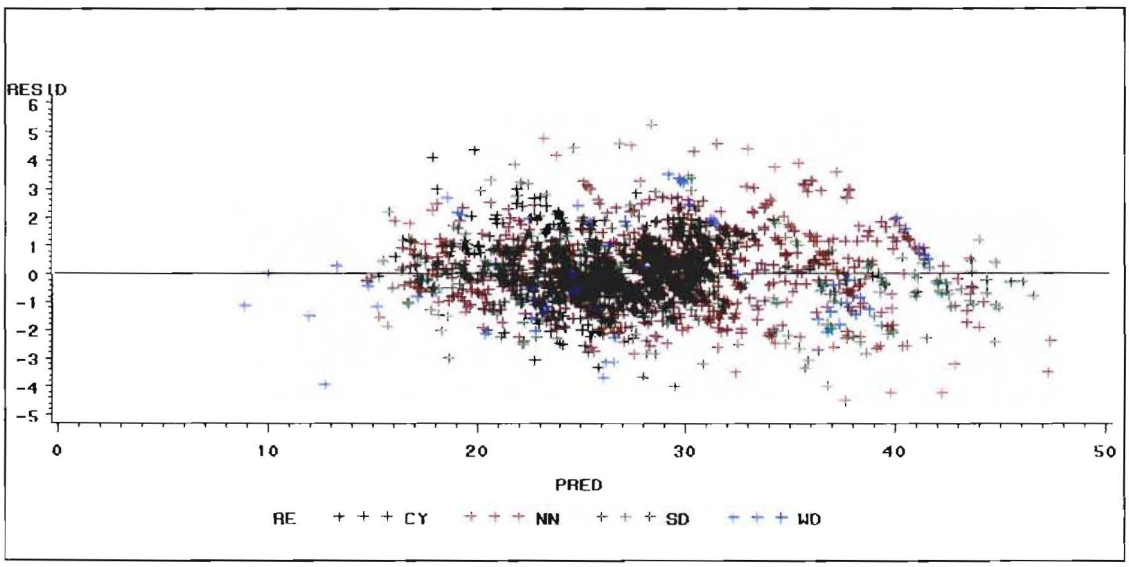

Figure 5.12: Plot of residuals vs predicted values when applying the best mixed interval equation for mean top height to the long interval data

5.3 Stem Survival/ha Equation

5.3.1 Applying the best short interval equation to the other data sets

A modified form of exponential equation 4.9 gave the best fit for projecting survival of stems/ha in the short interval data set:

$$N_2 = N_1 \exp\left((\alpha + \gamma K1)(T_2^{\beta} - T_1^{\beta})\right) \qquad (4.9)$$

When equation 4.9 was applied to the long and mixed interval data sets, the residual patterns were poor and greatly biased. The residuals showed that equation 4.9 underestimated mortality at low stems/ha in the long interval data set and underestimated mortality at low and high stems/ha in the mixed interval data set.

This equation had a mean residual error of 31.1878, while 95 percent of residuals were contained within ±262 stems/ha for the long interval data set; it had a mean residual error of 20.5895 and 95 percent of residuals lay within ±239 stems/ha for the mixed interval data set. The plots of residuals against predicted values are shown in Figures 5.13 and 5.14.

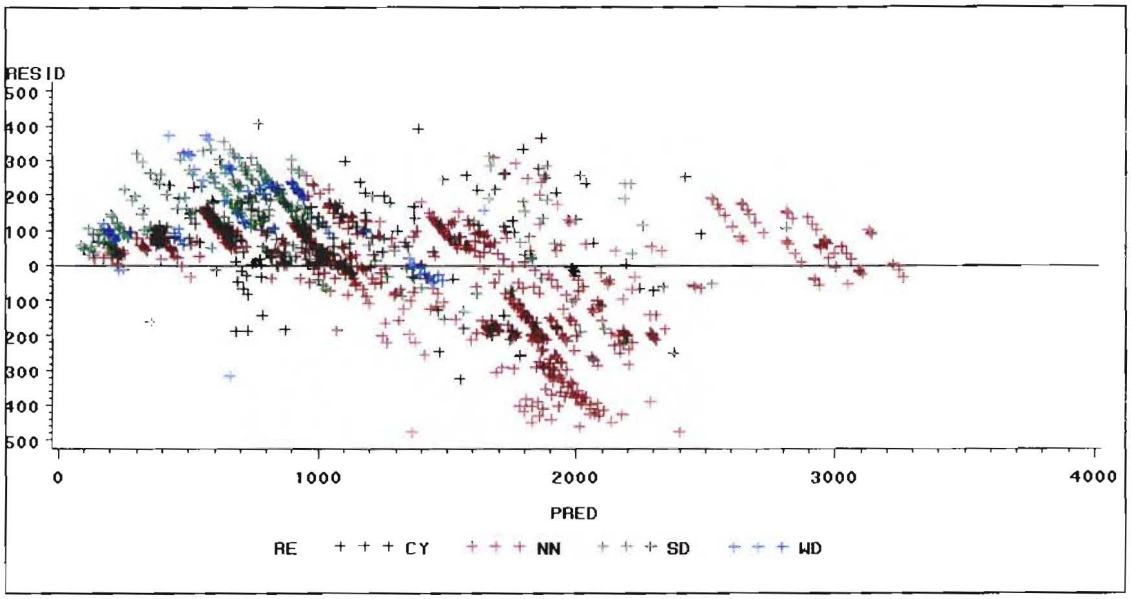

Figure 5.13: Plot of residuals vs predicted values when applying the best short interval equation for survival to the long interval data

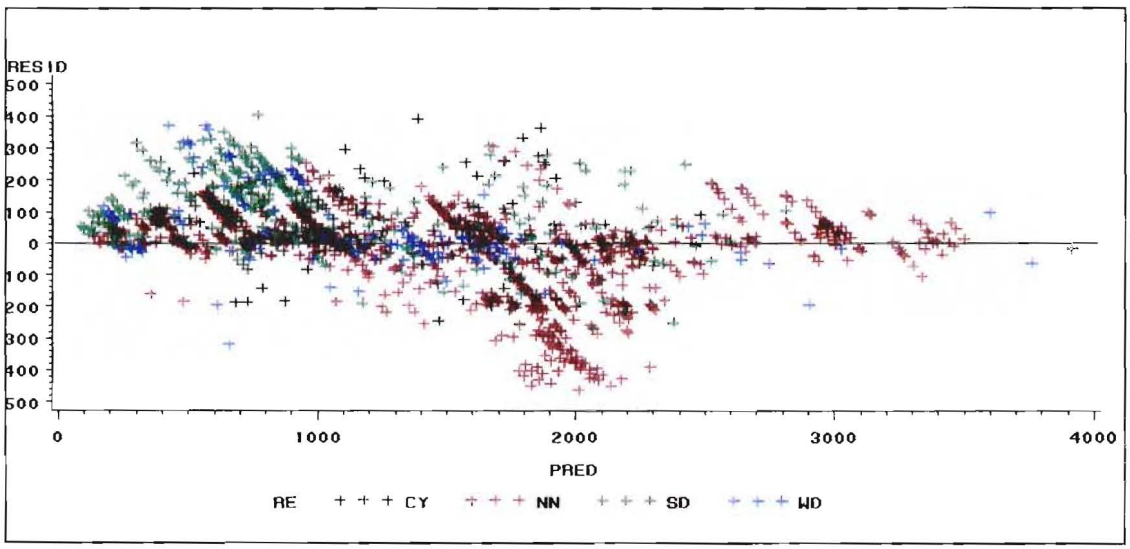

Figure 5.14: Plot of residuals vs predicted values when applying the best short interval equation for survival to the mixed interval data

5.3.2 Applying the best long interval equation to the other data sets

For the long interval data set, a modified Gompertz equation 4.15, which included site index and dummy variables, provided the best fit.

$$N_2 = N_1 e(\beta(T_2^2-T_1^2)) e(\alpha+\beta_1K_1+\beta_2K_2+\beta_3K_3+\beta_4S)(1-\exp(\beta(T_2^2-T_1^2))) \quad (4.15)$$

When equation 4.15, with its coefficients derived for the long interval data set, was applied to the short interval data set the residuals did not exhibit non-normality but tended to overestimate slightly. This equation had a mean residual error of –4.52, while 95 percent of residuals were contained within ±63 stems/ha.

When equation 4.15 was applied to the mixed interval data set, the residual pattern was nearly the same as that for the model chosen as best in the mixed interval data set. This similarity can be seen in Figure 5.16 and is a consequence largely of using the same functional form and similar coefficients. This equation had a mean residual error of 1.49, skewness of 0.073 and 95 percent of residuals lay within ±180 stems/ha, while the best mixed interval model had a mean residual error of 1.63, skewness of 0.07 and 95 percent of residuals lay within ±177 stems/ha. The plots of residuals against predicted values are shown in Figures 5.15 and 5.16.

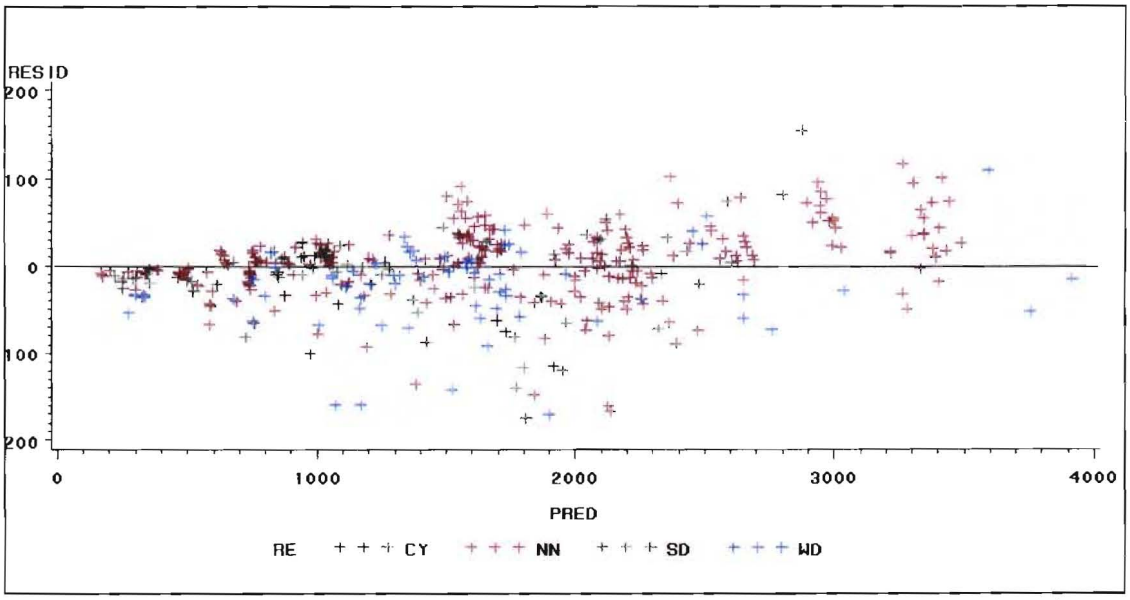

Figure 5.15: Plot of residuals vs predicted values when applying the best long interval for survival to the short interval data

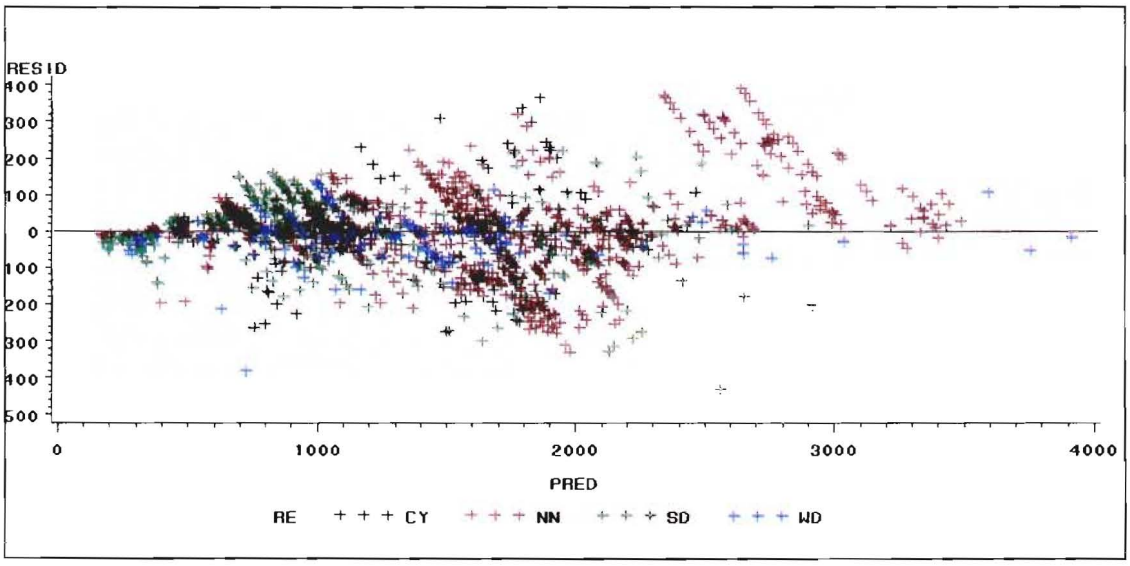

Figure 5.16: Plot of residuals vs predicted values when applying the best long interval equation for survival to the mixed interval data

5.3.3 Applying the best mixed interval equation to the other data sets

Mortality for the mixed interval data set was represented by a Gompertz functional form in 4.20.

$$N_2 = N_1 e^{(\beta(T_2^2-T_1^2))} e^{(\alpha+\beta_1K_1+\beta_2K_2+\beta_3K_3+\beta_4S)(1-\exp(\beta(T_2^2-T_1^2)))} \qquad (4.20)$$

When this equation with its coefficients was applied to the short interval data set, the result was the same as for the long interval equation, because of its similar coefficients. This equation had a mean residual error of –4.35, while 95 percent of residuals were contained within ±63 stems/ha.

When equation 4.20 was applied to the long interval data set the residual pattern was nearly the same as that for the model chosen as best in the long interval data set. This is shown in Figure 5.18. It is a consequence of having the same functional form and using similar coefficients. This equation had a mean residual error of 5.18, skewness of -0.031 and 95 percent of residuals lay within ±213 stems/ha, while the best model for the long interval had a mean residual error of 5.13, skewness of -0.010 and 95 percent of residuals lay within ±212 stems/ha. The plots of residual against predicted values are shown in Figures 5.17 and 5.18.

The applicability of the mixed interval data-based equations of basal area (G), mean to height (h_{100}) and stem survival/ha (N) to the other data sets showed good fits in general without apparent bias. But the applicability of the short interval data based equations was not as successful as those built using the mixed interval data. A summary of the results of cross fitting and with comparison of the original fits is in Table 5.5.

Table 5.5 Summary of statistics of cross fitting between the short and mixed equations

Basal area/ha equation				
Variable	Short best	Mixed to short	Mixed best	Short to mixed
Mean residuals	-0.10	-0.10	-0.06	0.28
Skewness	0.09	0.42	0.03	-0.29
95% residuals	±1.92	±2.42	±7.95	±9.89
Mean top height equation				
Variable	Short best	Mixed to short	Mixed best	Short to mixed
Mean residuals	0.03	-0.03	0.007	0.23
Skewness	0.35	0.36	0.30	0.71
95% residuals	±1.0	±1.0	±1.67	±2.10
Stem survival/ha equation				
Variable	Short best	Mixed to short	Mixed best	Short to mixed
Mean residuals	-0.67	-4.35	1.63	20.59
Skewness	-0.73	-0.62	0.07	-0.61
95% residuals	±53	±63	±177	±239

The mixed interval mean top height equation produced nearly the same values of mean residual error and skewness as those for the short data best model when applied to its data. The short interval model for basal area/ha and stem survival/ha equations did not produce as good a result for the mixed interval data set as shown Table 5.5. The mixed interval model, thus, proved to have superior compatibility compared with the other data sets.

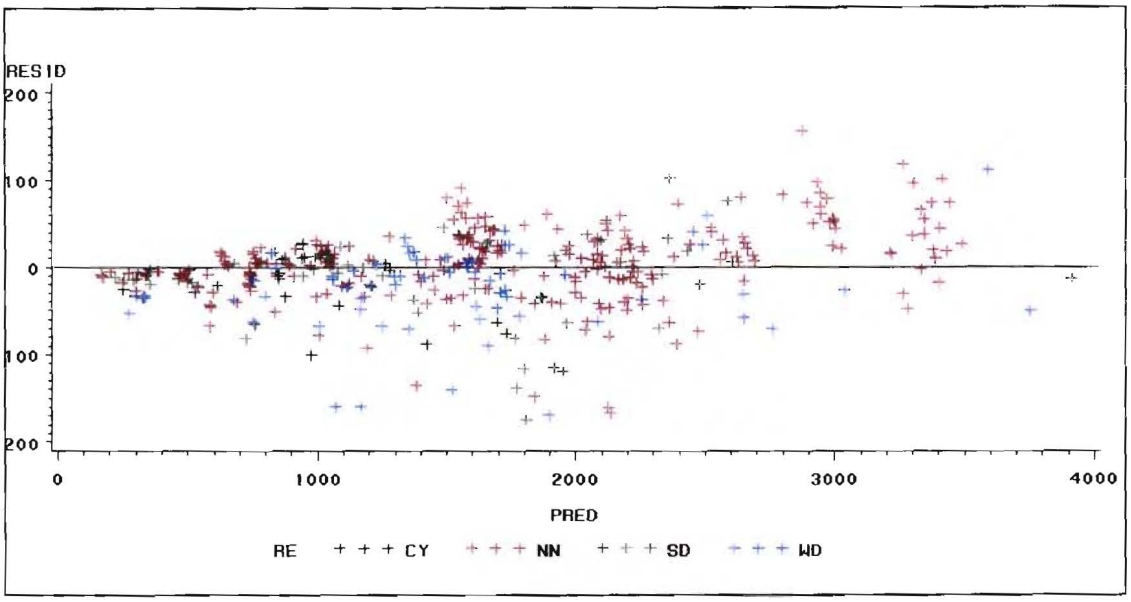

Figure 5.17: Plot of residuals vs predicted values when applying the best mixed interval equation for survival to the short interval data

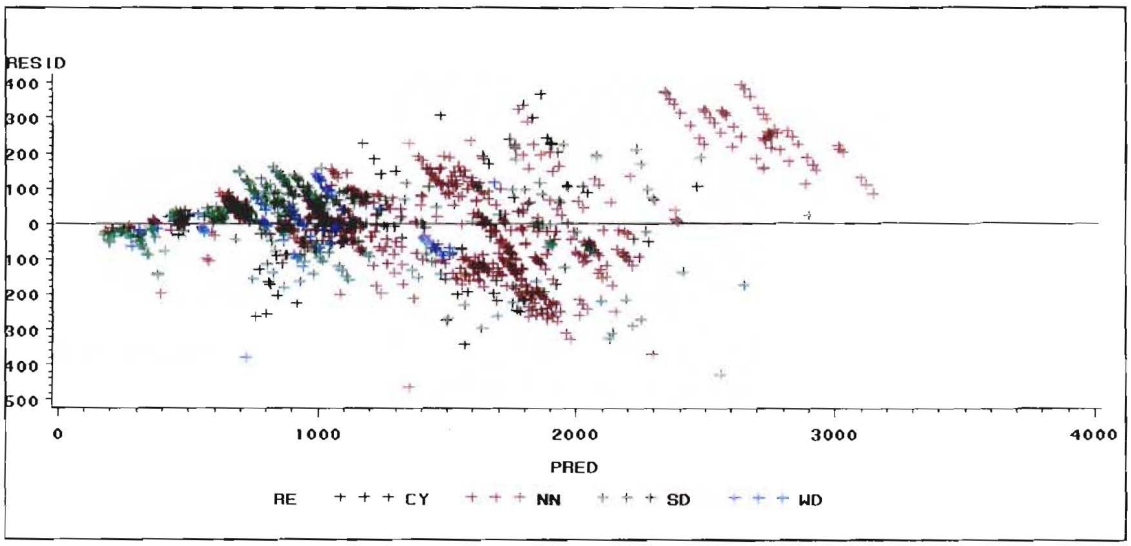

Figure 5.18: Plot of residuals vs predicted values when applying the best mixed interval equation for survival to the long interval data

Chapter 6

Growth Model with a Data Set Free from Auto-correlation

Permanent sample plot data obtained from repeated measurement on an individual plot contain correlated errors. That is contrary to the requirement in statistical analysis to have no auto-correlation. In practice, however, these correlations are frequently ignored and estimation of coefficients is assumed to be unbiased, although the hypothesis tests have no real validity. The aim of this component of the study was to examine the models developed using a reduced and less correlated data set, tested so as to ensure that regression coefficients were similar to the full set ones found earlier, and where significantly different from zero.

6.1 Data Set

One pair of measurements was selected randomly from each permanent plot to avoid auto-correlation. The data set consisted of only 366 observations, with a range of intervals from 1 to 28 years.

6.2 Net Basal Area Equation

Various growth functions of both anamorphic and polymorphic forms were fitted to the basal area/ha reduced data set. The same procedures which were used in

modelling the earlier full data sets were adopted for modelling this data set. The uncorrelated data used to build this new net basal area/ha equation are summarized in Table 6.1.

Table 6.1 A summary of data free from auto-correlation net basal area/ha equation

Region	N	Variables	Age (yrs)	G (m^2/ha)	Stems /ha	Site Index (m)	Alt (m)
Canterbury	53	Mean	30.3	38.8	974	26.47	371
		Minimum	10.0	0.8	74	16.40	150
		Maximum	57.0	79.3	2426	31.90	790
Nelson	145	Mean	28.1	45.4	1001	32.49	428
		Minimum	7.0	1.2	93	19.00	130
		Maximum	59.1	115.3	3533	41.00	625
Southland	122	Mean	32.6	49.9	900	29.37	210
		Minimum	9.7	6.8	148	19.00	50
		Maximum	78.0	105.8	3168	40.60	625
Westland	41	Mean	24.0	27.8	963	31.12	233
		Minimum	7.1	0.2	173	11.00	0
		Maximum	52.6	108	3704	37.40	330

The Schumacher polymorphic function as shown in equation 6.1 gave the best preliminary fit. It was then further refined and compared with the predictions from the full set.

$$G_2=G_1(T_1/T_2)^{\beta 1}e^{\alpha(1-(T_1/T_2)^{\beta 1})} \qquad (6.1)$$

Combinations of predictor variables such as site index, altitude, thinning index and dummy variables were then introduced to this Schumacher polymorphic basic form. Altitude and dummy variables applied independently to equation 6.1 gave worthwhile improvements to the basic model. Altitude was not included, however, in the final model, because it was found to have little effect when used in combination with dummy variables. The other predictor variables did not appear to improve

equation 6.1. Thus, equation 6.2 a Schumacher polymorphic form with three dummy explanatory variables was found best for this data set.

$$G_2=G_1(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+k_3\beta_3)(1-(T_1/T_2)^{\beta}) \quad (6.2)$$

Table 6.2 presents the successive improvement in which additional variables were introduced to the basic form of Schumacher polymorphic function and the coefficients estimation for equation 6.2 is shown in Table 6.3. Mean square error (MSE) of equation 6.2 was similar to that of the mixed interval equation (23.54) and higher than that of the short interval equation (1.45), while it was lower than the long interval equation (49.54).

Table 6.2 Subsequent improvement in fitting basal area equation

Input Variables	Error SS	% Reduction in ESS	MSE
Basic form (B)	13483.5901	-	37.4544
B, ALT	12505.7387	7.25	34.8349
B, Dummy(K1,K2,K3)	9468.2918	24.29	26.5963

Table 6.3 Coefficients for best basal area/ha equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	4.660594774	0.04951570386	9468.2918	361	26.5963
β	1.183942320	0.03184843192			
β1	0.243712321	0.04507471946			
β2	0.482215412	0.04879843095			
β3	0.506243450	0.06397004290			

The goodness of fit was evaluated through plots of residuals against predicted values and time as shown in Figures 6.1 and 6.2. The chart of residuals is shown in Figure 6.3.

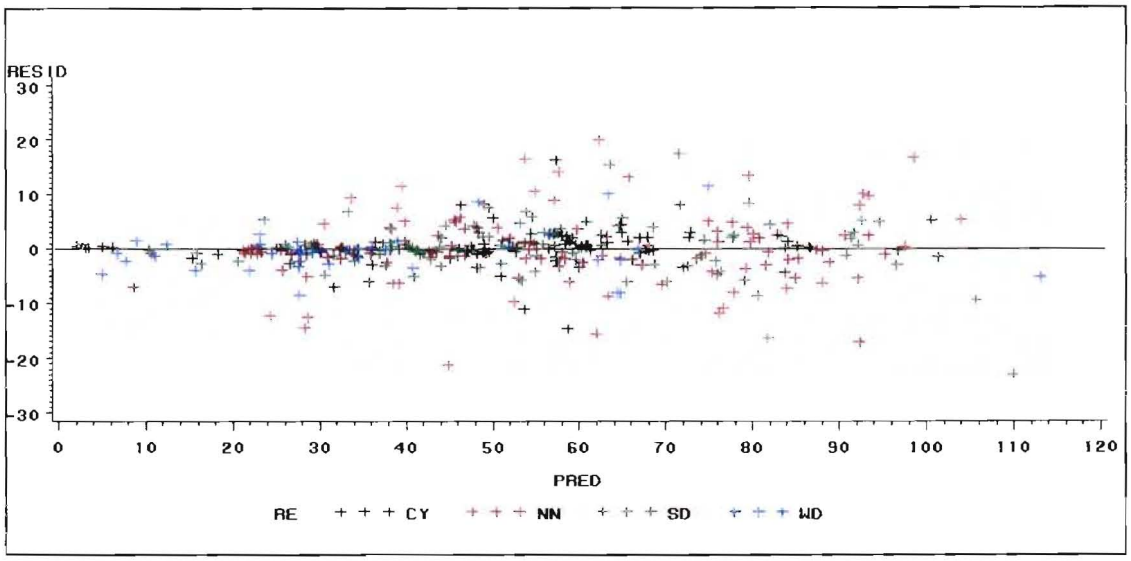

Figure 6.1: Plot of residuals vs predicted values for the basal area/ha equation from data free from auto-correlation

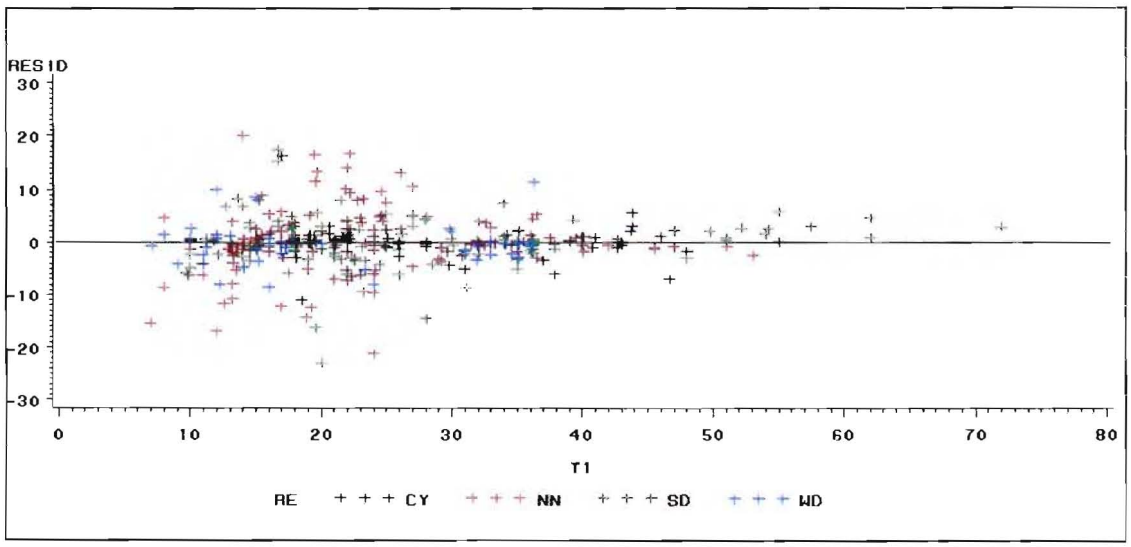

Figure 6.2: Plot of residuals vs age (T1 years) for the basal area/ha equation from data free from auto-correlation

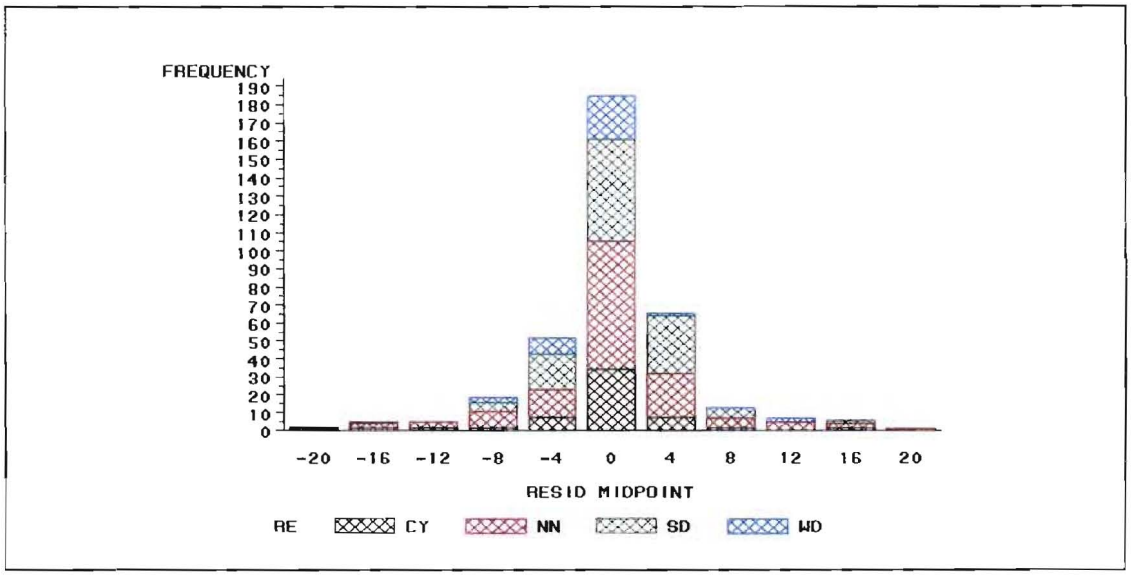

Figure 6.3: Frequency distribution of residuals for the basal area/ha equation from data free from auto-correlation

When equation 6.2 was fitted to the full short, long and mixed interval data for the same number of observations to compare the coefficients fairly, it showed that they were very similar, as shown in Table 6.4. When this equation was fitted with the short, long, mixed and data set free from auto-correlation using of the same degrees of freedom, the 95 percent confidence interval for the data set free from auto-correlation equation had similar degree of tightness with other data sets as shown Table 6.5. The t statistic, estimated by dividing the coefficient estimate by the asymptotic standard error, were greater for all coefficients included in the data set free from auto-correlation model, but smaller than those for long and mixed interval models.

Table 6.4 Comparison of coefficients for basal area/ha models among the four data sets

Parameter	Free correlated	Short	Long	Mixed
α	4.6605948	4.7622425	4.6875754	4.7147096
β	1.1839423	1.0935495	1.2325326	1.2234337
$\beta 1$	0.2437123	0.2863561	0.2142337	0.1964274
$\beta 2$	0.4822154	0.4250571	0.3543731	0.3295974
$\beta 3$	0.5062435	0.4292877	0.4409779	0.3979839

Table 6.5 Comparison of coefficients for net basal area/ha among four data sets, when the same number of observations as those of uncorrelated data set were used

Parameter	Estimates				Std. Error				95% Confidence Interval							
	Free from correlation	Short data	Long data	Mixed data	Independent data	Short data	Long data	Mixed data	Independent data		Short data		Long data		Mixed data	
									Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
α	4.6606	4.6759	4.6940	4.6454	0.0495	0.0647	0.0442	0.0400	4.563	4.758	4.549	4.803	4.607	4.781	4.567	4.724
β	1.1839	1.1606	1.2216	1.2809	0.0382	0.0420	0.0393	0.0385	1.109	1.259	1.078	1.243	1.144	1.299	1.205	1.357
β_1	0.2437	0.2196	0.2351	0.2351	0.0451	0.0602	0.0406	0.0362	0.155	0.332	0.101	0.338	0.155	0.315	0.164	0.306
β_2	0.4822	0.4136	0.4144	0.4000	0.0488	0.0587	0.0417	0.0372	0.386	0.578	0.298	0.529	0.332	0.496	0.327	0.473
β_3	0.5062	0.4504	0.4330	0.4081	0.0640	0.1097	0.0449	0.0602	0.380	0.632	0.235	0.666	0.345	0.521	0.290	0.527

6.3 Mean Top Height Equation

The data used for mean top height are summarized in Table 6.6. After trying several anamorphic and polymorphic of growth functional forms as was done when modelling for full sets of data, the Schumacher polymorphic equation 6.3 was found best

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta 1}e^{\alpha(1-(T_1/T_2)^{\beta 1})} \quad (6.3)$$

Table 6.6 A summary of data for mean top height equation

Region	No.	Variable	Mean	Minimum	Maximum
Canterbury	53	Age(yrs)	30.3	10.0	57.0
		h_{100} (m)	20.8	3.1	37.1
Nelson	145	Age(yrs)	28.1	7.0	59.1
		h_{100}(m)	23.6	5.6	44.1
Southland	124	Age(yrs)	32.4	9.7	78.0
		h_{100} (m)	24.1	5.8	45.8
Westland	43	Age(yrs)	25.1	7.1	59.1
		h_{100} (m)	17.8	3.3	42.1

Various of modifications were introduced to equation 6.3 using additional explanatory variables. Equation 6.4 with three dummy variables for regions was found best.

$$h_{100,2}=h_{100,2}(T_1/T_2)^{\beta}e^{(\alpha+K_1\beta_1+K_2\beta_2+k_3\beta_3)(1-(T_1/T_2)^{\beta})} \quad (6.4)$$

Equation 6.4 is of the same form as was found best in the long and the mixed interval data sets, and of similar form to the best model for the short interval data with only one dummy variable. The coefficients for this equation are shown in Table 6.7. The mean square error of equation 6.4 was higher than that of each full data sets equations: 0.32 for the short interval equation; 1.72 for the long interval equation; and 1.04 for the mixed interval equation. The goodness of fit was evaluated through plots of residuals against predicted values and time as shown in Figure 6.4 and Figure 6.5. The chart of residuals is shown in Figure 6.6.

Table 6.7 Coefficients for mean top height equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	5.360334863	0.18729318281	598.03211	349	1.73847
β	0.377657081	0.02745517962			
$\beta 1$	0.432666670	0.07315013651			
$\beta 2$	0.270921645	0.07836580040			
$\beta 3$	0.393372903	0.09912445828			

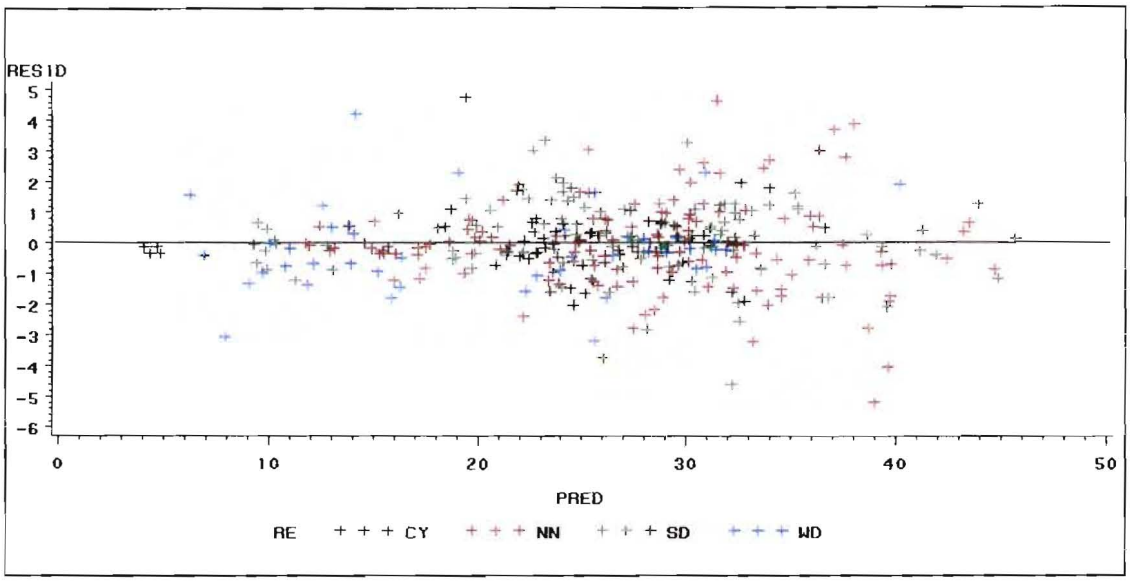

Figure 6.4: Plot of residuals vs predicted values for mean top height equation from data free from auto-correlation

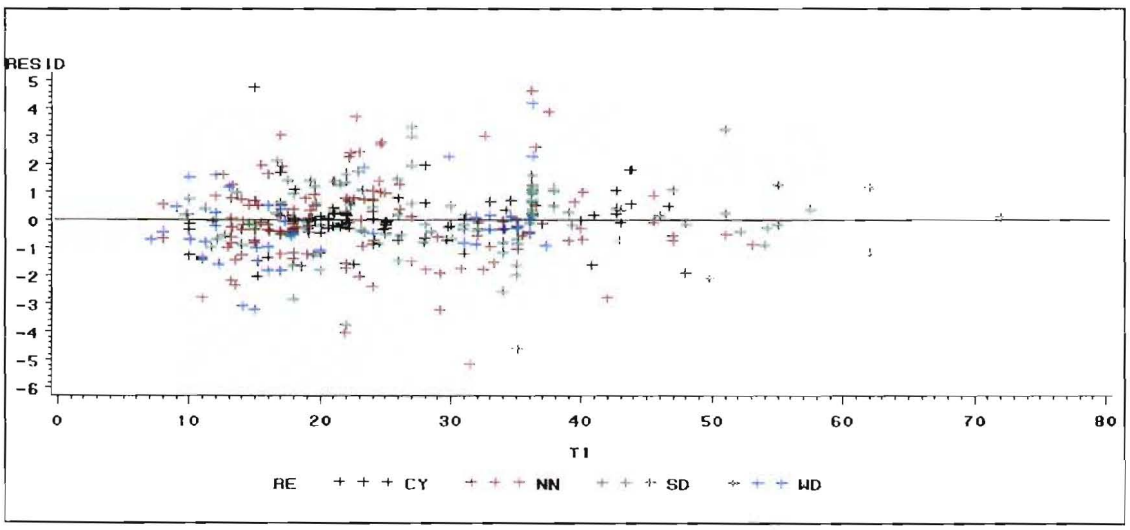

Figure 6.5: Plot of residuals vs age (T1 years) for mean top height equation from data free from auto-correlation

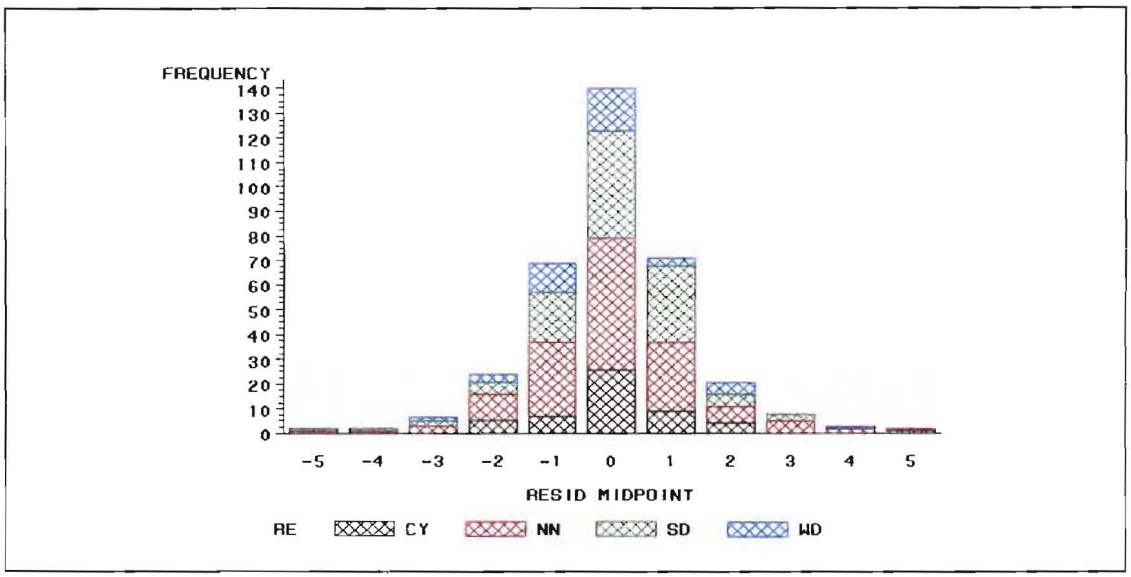

Figure 6.6: Frequency distribution of residuals for mean top height equation from data free from auto-correlation

Equation 6.4 was fitted to two data sets: one consisted of the same number of observations for the short, the long and the mixed interval data: the other contained the same number of observation as those in the data set free from auto-correlation to compare the coefficients. They were found to be very closely related, as shown in Tables 6.8 and 6.9. Coefficients of dummy variables $\beta 1$, $\beta 2$ and $\beta 3$ representing regions, however, were not significant at $\alpha = 0.05$ level in both short interval data sets unlike best short interval mean top height equation for full data set.

Table 6.8 Comparison of coefficients for mean top height among the four data sets

Parameter	Free from correlated	Short	Long	Mixed
α	5.3603349	5.1226059	5.4831974	5.3825376
β	0.3776471	0.4418125	0.3701572	0.3908114
$\beta 1$	0.4326667	-	0.3787746	0.3123980
$\beta 2$	0.2709216	-	0.2074681	0.1692655
$\beta 3$	0.3933729	-	0.4378491	0.3746830

Table 6.9 Comparison of coefficients for mean top height among four data sets, when the same number of observations as those of uncorrelated data set were used

Parameter	Estimates				Std. Error				95% Confidence Interval							
	Free from correlation	Short data	Long data	Mixed data	Independent data	Short data	Long data	Mixed data	Independent data		Short data		Long data		Mixed data	
									Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
α	5.3603	5.5706	5.3230	5.8314	0.1873	0.3712	0.1624	0.1884	4.992	5.729	4.841	6.301	5.004	5.642	5.461	6.202
β	0.3776	0.3702	0.3937	0.3330	0.0275	0.0487	0.0250	0.0205	0.324	0.432	0.274	0.466	0.345	0.443	0.293	0.373
β_1	0.4327	0.2727	0.3400	0.3248	0.0732	0.1756	0.0616	0.0686	0.289	0.577	-0.07	0.618	0.219	0.461	0.190	0.460
β_2	0.2709	0.1913	0.1265	0.1953	0.0784	0.1811	0.0634	0.0679	0.117	0.425	-0.17	0.548	0.002	0.251	0.062	0.329
β_3	0.3934	0.1569	0.3954	0.3550	0.0991	0.2340	0.0705	0.0999	0.198	0.588	-0.30	0.617	0.257	0.534	0.158	0.552

6.4 Stem Survival/ha Equation

The data used for the stem survival/ha equation were included only when mortality occurred over an interval, just as had been done for the whole data sets. The data consisted of 168 observations and are summarized in Table 6.10.

Table 6.10 Summary data for stem survival/ha equation

Region	No.	Variable	Mean	Minimum	Maximum
Canterbury	29	Age(yrs)	31.7	15.0	57.0
		Stems/ha	1254	350	2426
Nelson	68	Age(yrs)	29.7	7.0	59.0
		Stem/ha	1344	93	3533
Southland	48	Age(yrs)	28.8	9.7	64.0
		Stem/ha	1357	148	2921
Westland	23	Age(yrs)	22.7	7.1	59.1
		Stem/ha	1412	173	3704

A range of anamorphic and polymorphic survival functions was fitted to this reduced, uncorrelated data set. An anamorphic form of exponential decay function, 6.5, was adopted after it proved to fit the data best.

$$N_2 = N_1 \exp(\alpha\,(T_2 - T_1)) \tag{6.5}$$

Including site index did not make any improvement but dummy variables for the regions did substantially improve the equation. Therefore, equation 6.6 was found to be the best fitting equation.

$$N_2 = N_1 \exp((\alpha + \beta_1 k_1 + \beta_2 k_2 + \beta_3 k_3)(T_2 - T_1)) \tag{6.6}$$

The graph of the residuals lay mainly within ±250 trees per ha, as shown in Figure 6.7. The plots of the residuals against age and frequency distribution chart are given

in Figures 6.8 and 6.9. Table 6.11 shows the coefficients for the stem survival/ha equation derived from data set free from auto-correlation.

Table 6. 11 Coefficients for stem survival/ha equation

Parameter	Estimates	Std. Error	Error SS	N	MSE
α	-0.02392743	0.00199552480	3812910	168	23249.45
β2	0.007443679	0.00221589948			
β2	0.008260442	0.00261816604			
β3	0.012047764	0.00302173990			

However, it was not impossible to compare the coefficients of this equation with those of the short, the long and the mixed interval data sets, because the best model in each data set had very different functional forms. Consequently, equation 6.7 was used for comparative purposes to fit all four data sets.

$$N_2=N_1e^{(\alpha(T_2-T_1))} \qquad (6.7)$$

When equation 6.7 was fitted to the short, long and mixed interval data, it showed similar coefficients. Table 6.12 shows the comparison of coefficients.

Table 6.12 Comparison of coefficients for mortality equation with four data

Parameter	Free correlated	Short	Long	Mixed
α	-0.0174038	-0.0177928	-0.0165937	-0.0175088

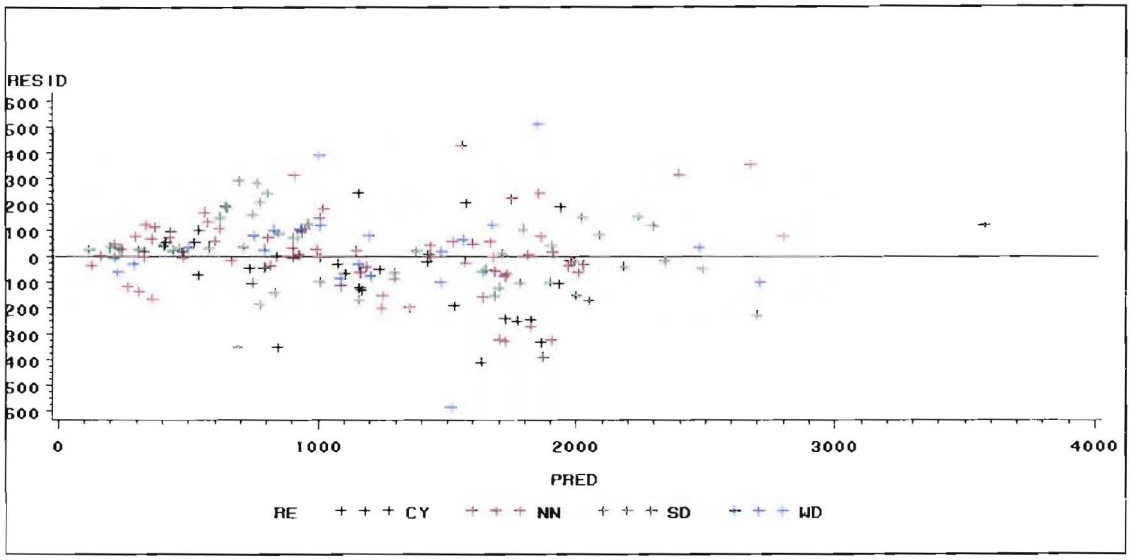

Figure 6.7: Plot of residuals vs predicted values for stem survival/ha equation from data free from auto-correlation

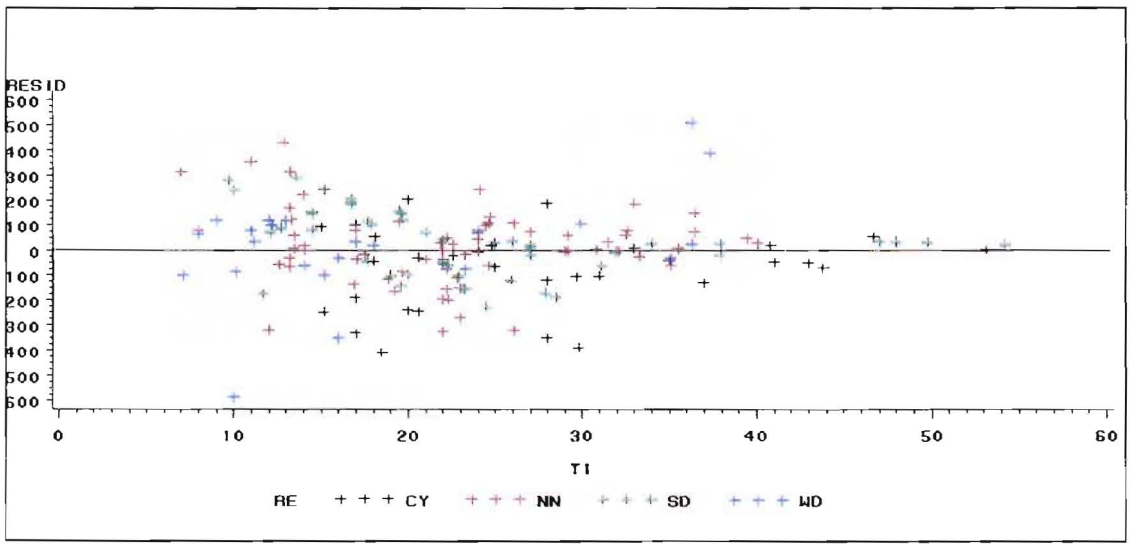

Figure 6.8: Plot of residuals vs age (T1 years) for stem survival/ha equation from data free from auto-correlation

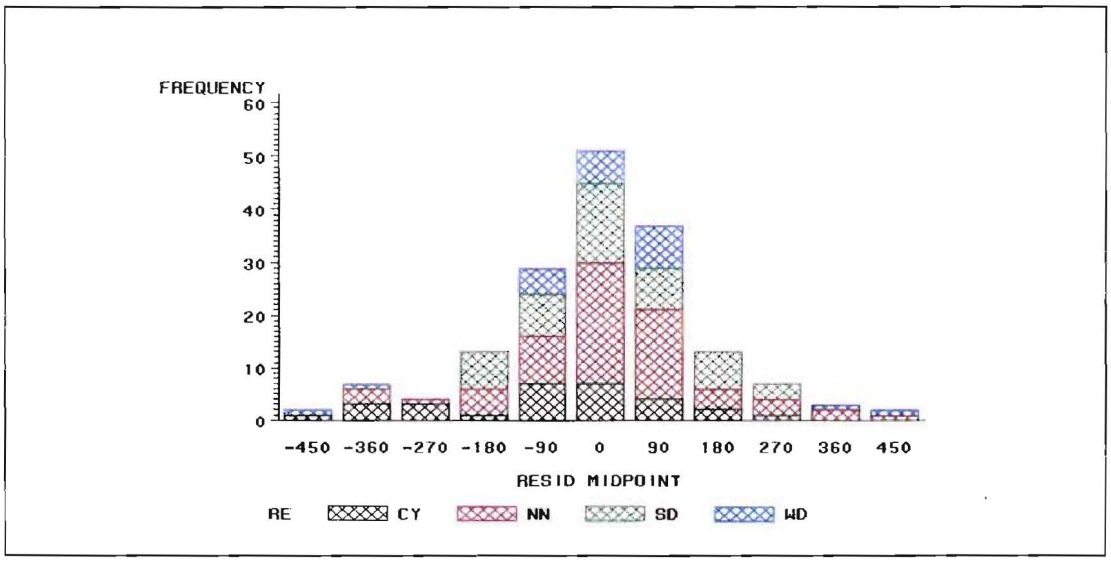

Figure 6.9: Frequency distribution of residuals for stem survival/ha equation from data free from auto-correlation

Chapter 7

Evaluation of Growth Model Predictions

7.1 Verification and Validation

Any growth and yield equation should be logically consistent and biologically realistic. Verification involves an examination of the structure and properties of a model, with or without supplementary data, to confirm that it has no internal inconsistencies and it is biologically realistic (Bruce and Wensel, 1987; Olderwald and Hans, 1993; Vanclay and Skovsgaard, 1996). The soundness and reliability of the predictions made in this study were accomplished with graphical analyses as well as statistical indices. These were performed by way of residual patterns, frequency distribution of residuals and ensuring that coefficient estimates of all equations differed significantly from zero. The residual plots and PROC UNIVARIATE procedures in SAS (SAS Inc, 1990) were used for visual and numerical interpretations of residual patterns and statistics.

The process of validation attempts to increase confidence in a model's ability to provide useful and correct inferences about the growth of stands, rather than to prove that a model is correct. It needs to provide valuable information to the developer in formulating and improving subsequent models (Newberry and Stage, 1987; Goulding, 1979). The most convincing validation is done using an independent data

set. Vanclay (1994) emphasized the need for independent data to be spatially (e.g. different location), temporally (e.g. more recent), or logistically (e.g. controlled by a different agency) independent. Such data sets, however, are rarely available as is the case for Douglas-fir in the South Island. The other method of determining how well a model will perform is to compare predictions from the model with those of an existing model.

7.1.1 Residual Patterns and Frequency Charts

Plotting of the residuals from each equation against predicted values and other independent variables such as age and altitude provides valuable insights into the nature of bias of equations. Normally distributed residuals should exist uniformly along the zero reference line, while biased residuals would show some systematic departure from such a distribution.

The forms of bar charts for frequency distribution of residuals were also used to judge normality of distributions. These charts, for the whole of the South Island and for individual regions, were examined by plotting residual frequencies against residual midpoints. The results of analysing of residual patterns and frequency distributions showed no serious bias for basal area/ha and mean top height projection equations for the three different interval length data sets analysed here. The residual bar chart involves an arbitrary choice to class widths. Consequently, other statistics were given greater credence when checking the goodness of fit more objectively.

7.1.2 Other Statistics

The standard deviation of residuals used to provide a dispersion of residuals from their mean, which were computed with random normal deviate (RND) in SAS.

RND = (ABS (RESID)/RMS)

Where ABS(RESID) is an absolute value of the residual for each observation and RMS is the root mean square of residuals. When the observations were more than 3.5 times the RND, they were regarded as outliers and traced back for further examination.

The Univariate procedure in SAS was also used to provide residual statistics for quantitative and qualitative analyses of residuals. For each equation, the mean of residual, variance, kurtosis, skewness and the extreme value were carefully assessed.

(1) Mean of Residual Error

The mean and sum of residual errors should, theoretically, be equal to zero. In reality, however, these values for equations are rarely exactly zero and are usually greater or less than this amount when equations are fitted with non-linear least-squares estimation procedures. They should, however, be close to zero. The value of the mean residual is more important than its sign which shows that an equation tends to underestimation or overestimation in terms of judging on accuracy of the functions. The absolute mean residual, sum of absolute residuals divided by the sample size was used, as a measure of the departure from the true values. This is a measure of the average error prediction of the equation and should be as low as possible.

(2) Kurtosis and Skewness

The values of skewness and kurtosis were examined to ascertain the normality of distribution of residuals. Skewness of the normal distribution should be zero, while kurtosis values should lie between –2 and positive infinity. Equations with outside these range values of kurtosis were often found to have outliers.

(3) Extreme Values of Residuals

Extreme values (minimum and maximum) of residuals were used to check the precision of the equations. Percentiles were also used as indicators for practical limits to the variance.

7.1.3 Comparisons of Models

Two previous models exist for Douglas-fir grown throughout the South Island of New Zealand, namely SIDFIR developed by Law (1990) and DirfirStand developed by Temu (1992). The residual patterns and statistics for those existing models were compared to those of the new model after applying them to three different data sets. In conducting this comparison, useful insights could be obtained because the data used for development of these models came from the same regions and the same PSP data set. The values of the statistics are presented in Table 7.1.

Table 7.1 Comparison of residual statistics for new model, SIDFIR and DirfirStand

Data	Equation	Model	Residual statistics					
			N	RMS	Mean	ABS	Min	Max
Short interval	Basal area (m^2/ha)	New	1932	1.241	-0.009	0.813	-4.00	3.92
		DIRFIR	1932	-	-0.018	0.848	-4.64	4.39
		SIDFIR	1932	-	-0.218	2.635	-32.05	22.7
	Mean top height (m)	New	1825	0.309	0.030	0.424	-1.81	1.94
		DIRFIR	1825	-	0.042	0.448	-1.93	2.13
		SIDFIR	1825	-	0.243	0.661	-3.9	3.6
	Mortality (n/ha)	New	446	1385.8	-0.674	27.6	-142	101
		DIRFIR	446	-	-2.134	29.9	-158	147
		SIDFIR	446	-	9.758	91.4	-899	512
Long interval	Basal area (m^2/ha)	New	1780	43.37	-0.011	5.121	-21.7	22.0
		DIRFIR	1780	-	0.118	5.779	-32.84	23.43
		SIDFIR	1780	-	1.354	6.843	-27.24	36.2
	Mean top height (m)	New	1696	1.697	0.023	1.008	-4.51	4.59
		DIRFIR	1696	-	0.438	1.304	-5.65	5.67
		SIDFIR	1696	-	0.311	1.026	-3.9	3.6
	Mortality (n/ha)	New	1078	14512.4	5.127	90.6	-431	390
		DIRFIR	1078	-	4.190	96.0	-455	360
		SIDFIR	1078	-	-4.637	115.0	-592	512
Mixed interval	Basal area (m^2/ha)	New	3691	18.731	-0.065	2.821	-16.25	15.73
		DIRFIR	3691	-	0.083	3.079	-23.53	20.87
		SIDFIR	3691	-	0.534	4.504	-32.05	24.84
	Mean top height (m)	New	3506	0.881	0.007	0.693	-3.36	3.33
		DIRFIR	3506	-	0.244	0.836	-4.87	4.80
		SIDFIR	3506	-	0.282	0.824	-3.9	3.6
	Mortality (n/ha)	New	1533	10544.3	1.631	72.9	-426	374
		DIRFIR	1533	-	0.811	76.7	-454	353
		SIDFIR	1533	-	-1.573	107.3	-899	512

Where

N = number of re-measurements,

RMS = residual mean square,

Mean = mean residual,

ABS = mean of absolute residual,

Min = minimum residual value,

Max = maximum residual value.

(1) Net Basal Area/ha Equation

The net basal area/ha model developed in this study for the short interval data had a mean residual of -0.009 m^2/ha, which represents a slight overprediction and an absolute residual of 0.81 m^2/ha, which means that the equation would predict net basal area/ha with an average error of 0.81 m^2/ha. This represents a more than 30 % improvement compared with the absolute error of the SIDFIR at 2.64 m^2/ha. The new equation was also much less biased and more precise (Figure 7.1) compared with the SIDFIR model (Figure 7.2) and slightly better than DirfirStand model (Figure 7.3).

The net basal area/ha for the long interval data had a mean residual of -0.011 m^2/ha, indicating that the equation slightly over-predicts net basal area/ha. The estimates had an average deviation of 5.12 m^2/ha as compared to 5.78 m^2/ha for DifirStand and 6.83 m^2/ha for SIDFIR, meaning that a more precise estimation was accomplished with the new model. The residual patterns are shown in Figures 7.4, 7.5 and 7.6.

The mean residual and absolute mean residual values for net basal area/ha with the mixed interval data were -0.065 and 2.82 m^2/ha respectably, showing that slightly over-predicted basal area/ha with average deviation 2.82 m^2/ha. These values were slightly better than those for DirfirStand and much better than those for SIDFIR. The highest error was 15.73 m^2/ha and the lowest error was –16.25 for the new model, while the highest error in SIDFIR was 24.84 m^2/ha and the lowest was –32.05 m^2/ha. The residual patterns are shown in Figures 7.7, 7.8 and 7.9.

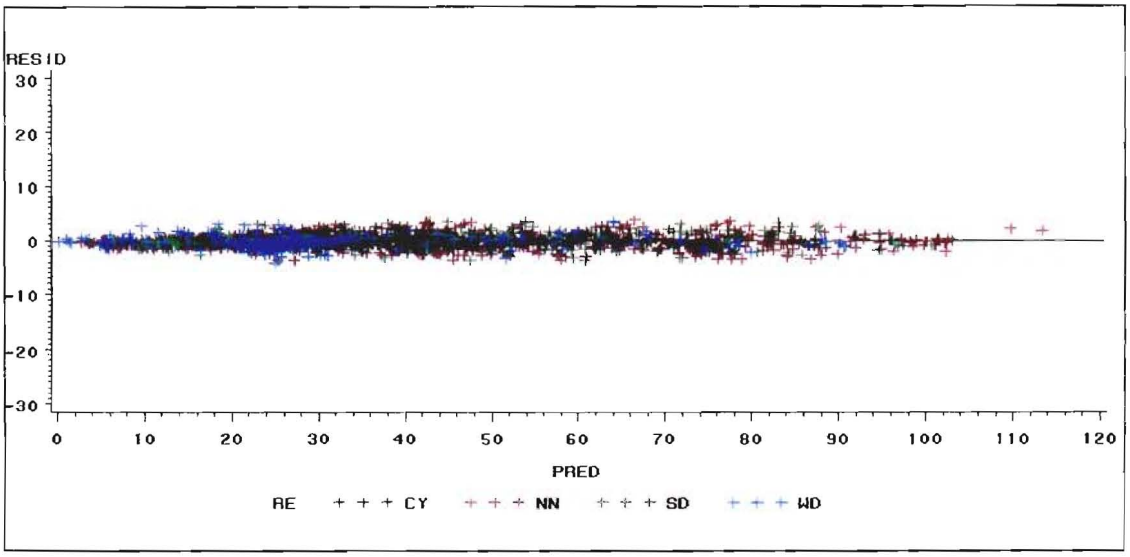

Figure 7.1: Plot of residuals vs predicted value for the short interval basal area/ha equation with new model

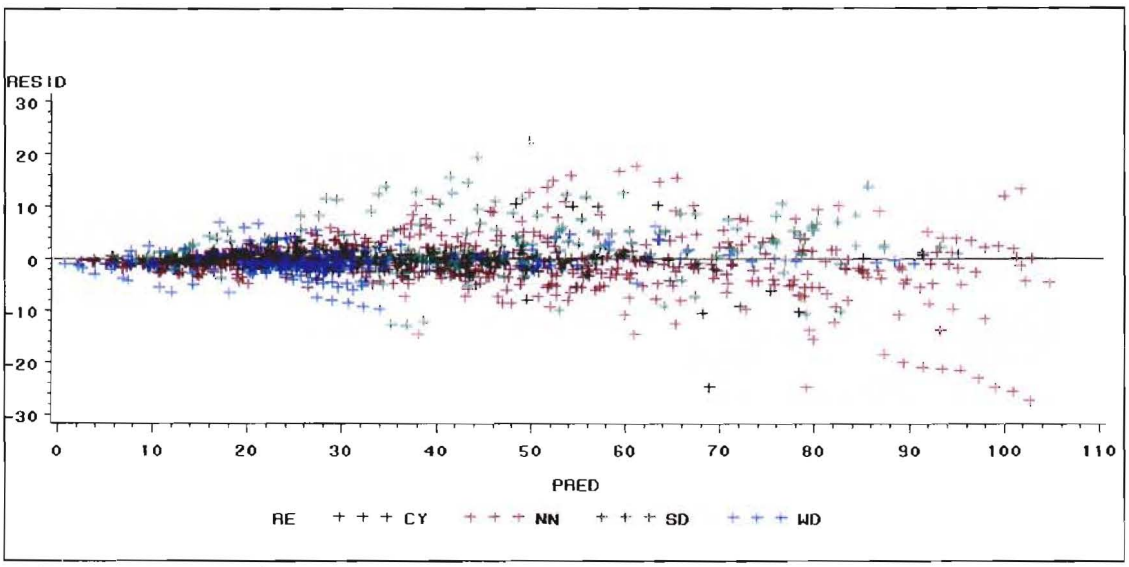

Figure 7.2: Plot of residuals vs predicted values for the short interval basal area/ha equation with SIDFIR

Lahti (1990)

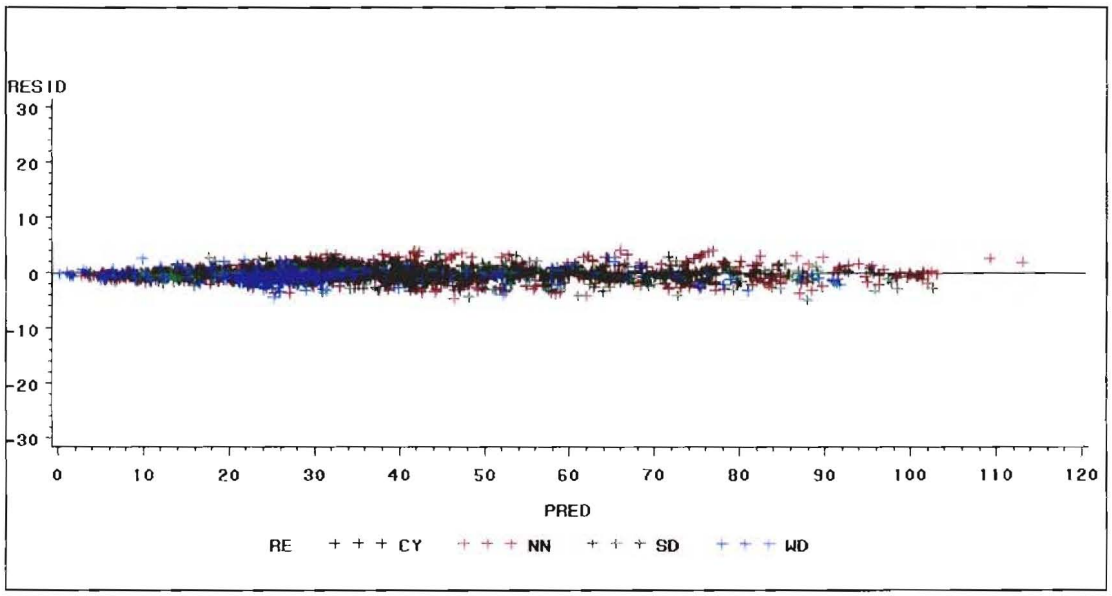

Figure 7.3: Plot of residuals vs predicted values for the short interval basal area/ha equation with DfirStand

TEMA (1997)

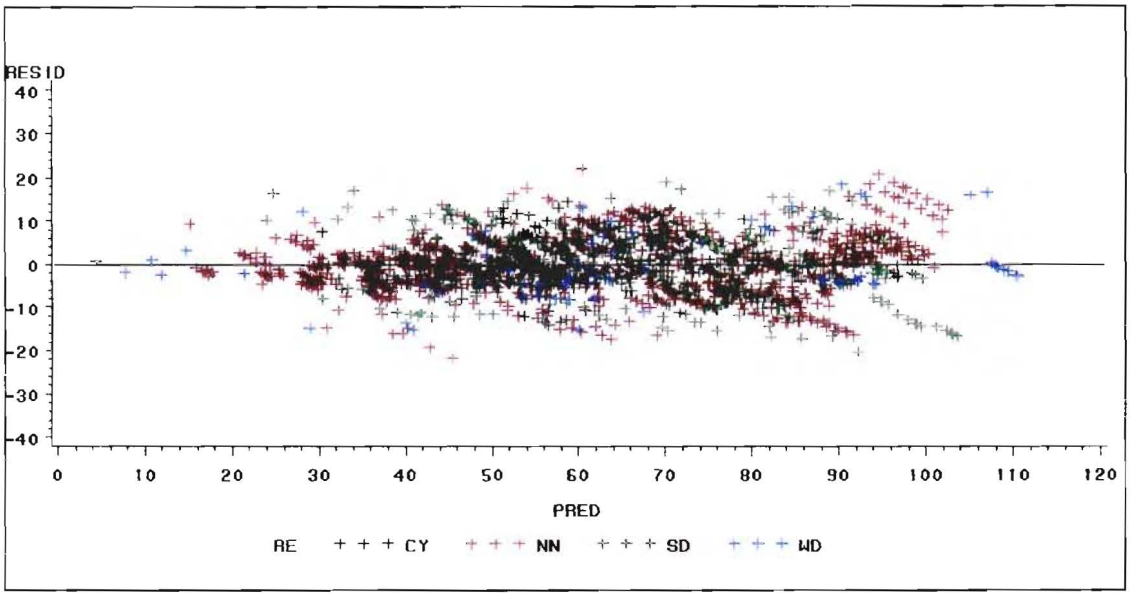

Figure 7.4: Plot of residuals vs predicted values for the long interval basal area/ha equation with new model

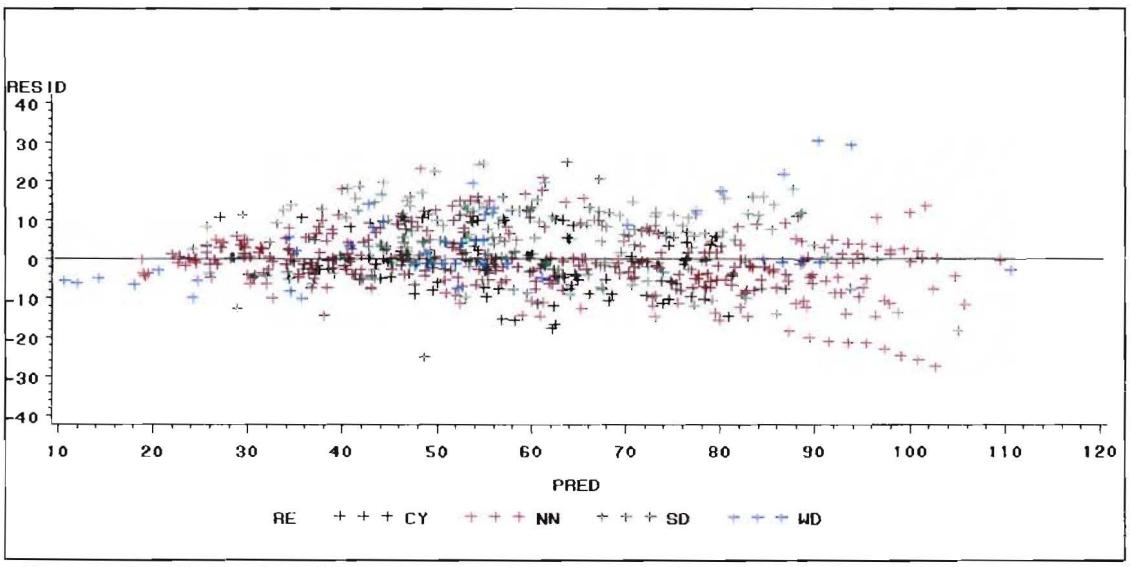

Figure 7.5: Plot of residuals vs predicted values for the long interval basal area/ha equation with SIDFIR

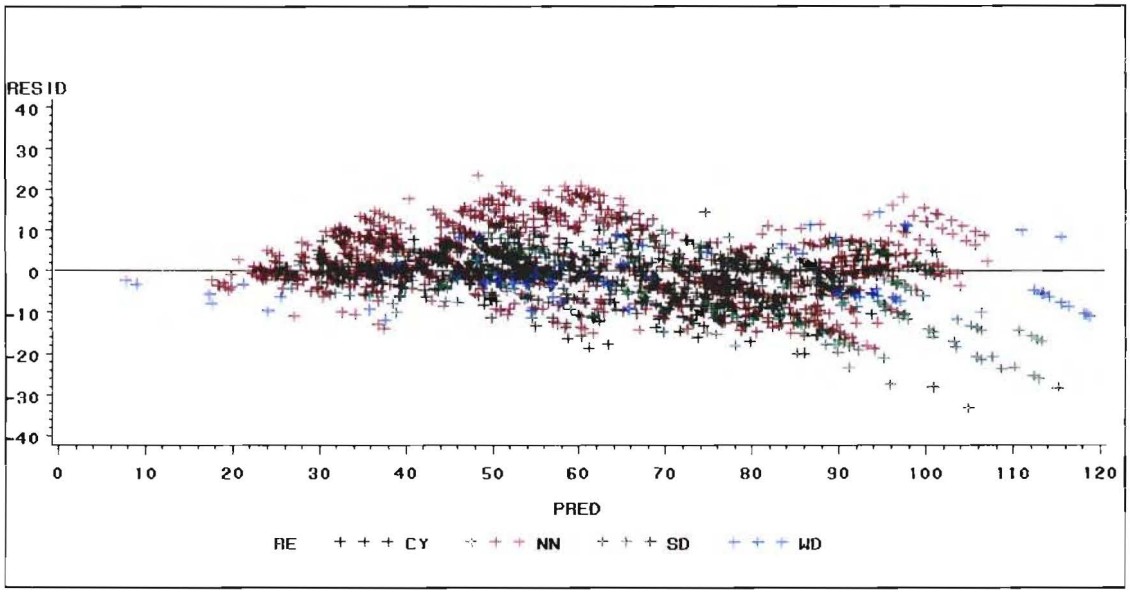

Figure 7.6: Plot of residuals vs predicted values for the long interval basal area/ha equation with DfirStand

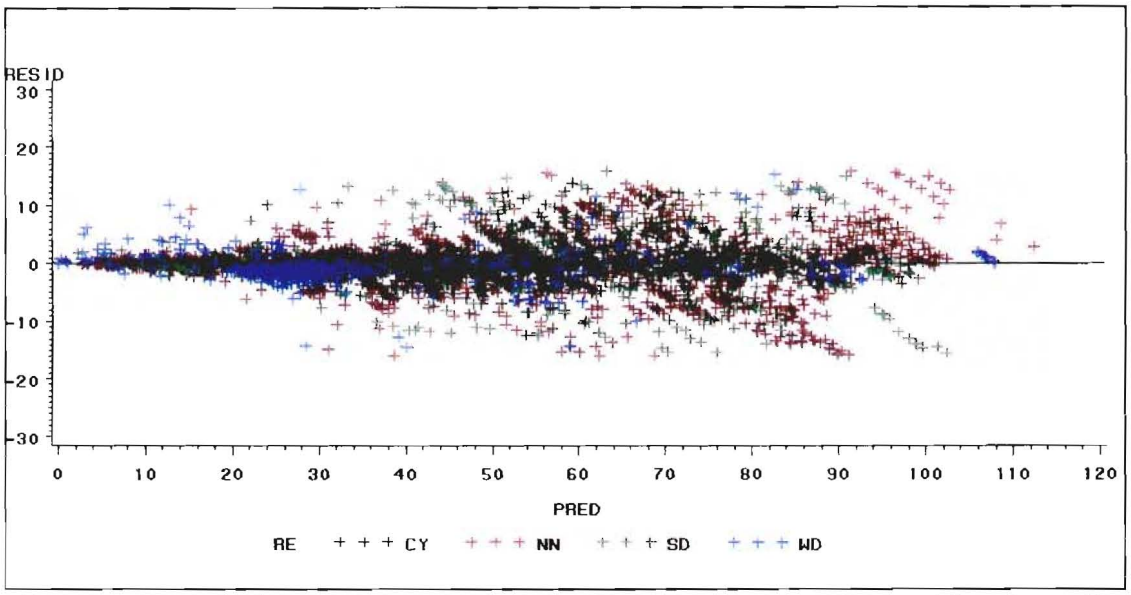

Figure 7.7: Plot of residuals vs predicted values for the mixed interval basal area/ha equation with new model

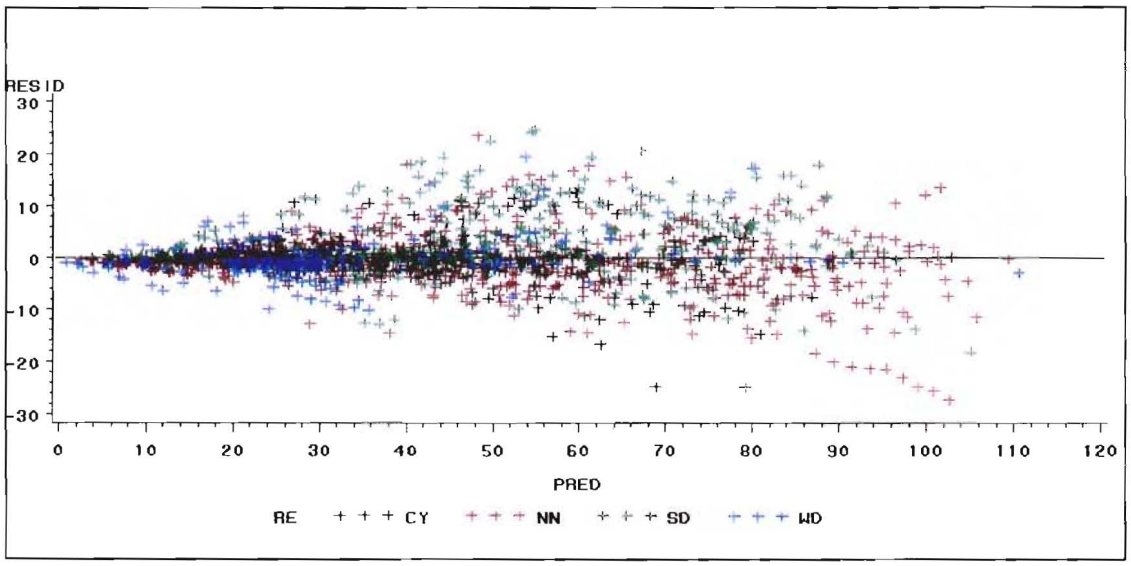

Figure 7.8: Plot of residuals vs predicted values for the mixed interval basal area/ha equation with SIDFIR

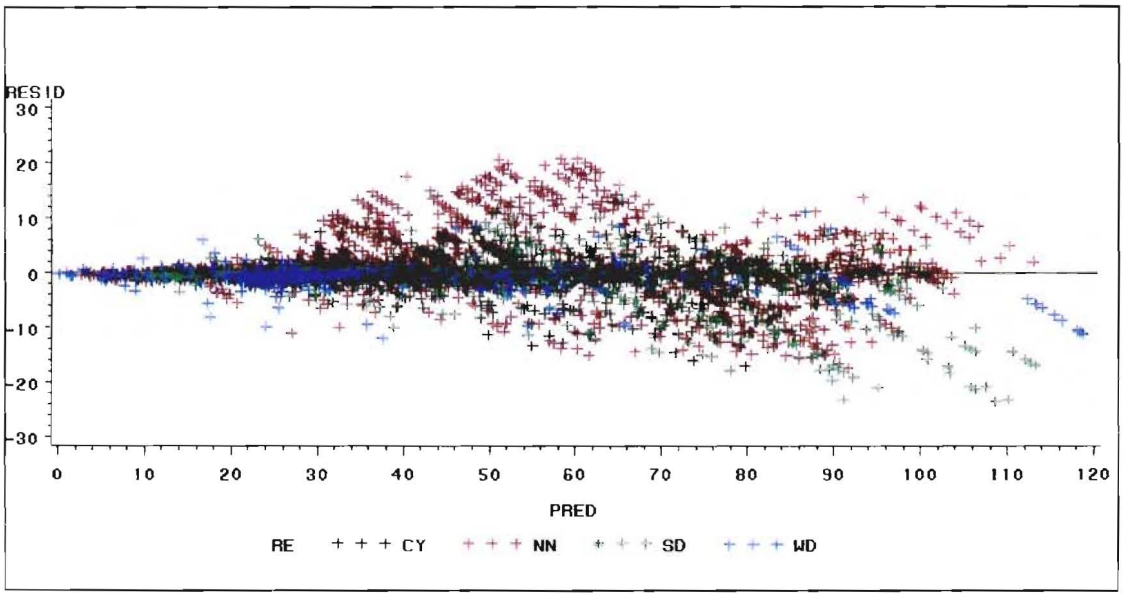

Figure 7.9: Plot of residuals vs predicted values for the mixed interval basal area/ha equation with DfirStand

(2) Mean Top Height Equation

Table 7.1 shows that all models provide estimates of height growth which are generally acceptable, with the average error within 0.5 m of the observed value. The mean top height equations of the new model for the short, long and mixed data slightly under-predict tree height by an average deviations 0.42, 1.01 and 0.69 m respectively. These values are better than those for SIDFIR which were 0.66, 1.03 and 0.82 respectively, and for DirfirStand were 0.45, 1.30 and 0.84 respectively. There are about 20 and 10 percent improvements for the short and the mixed data compared with the SIDFIR model. The new equations developed in this study were less biased compared with the SIDFIR and DfirStand models and gave more precise estimates for tree height in each data set. The residual patterns are shown in Figures 7.10 to 7.18.

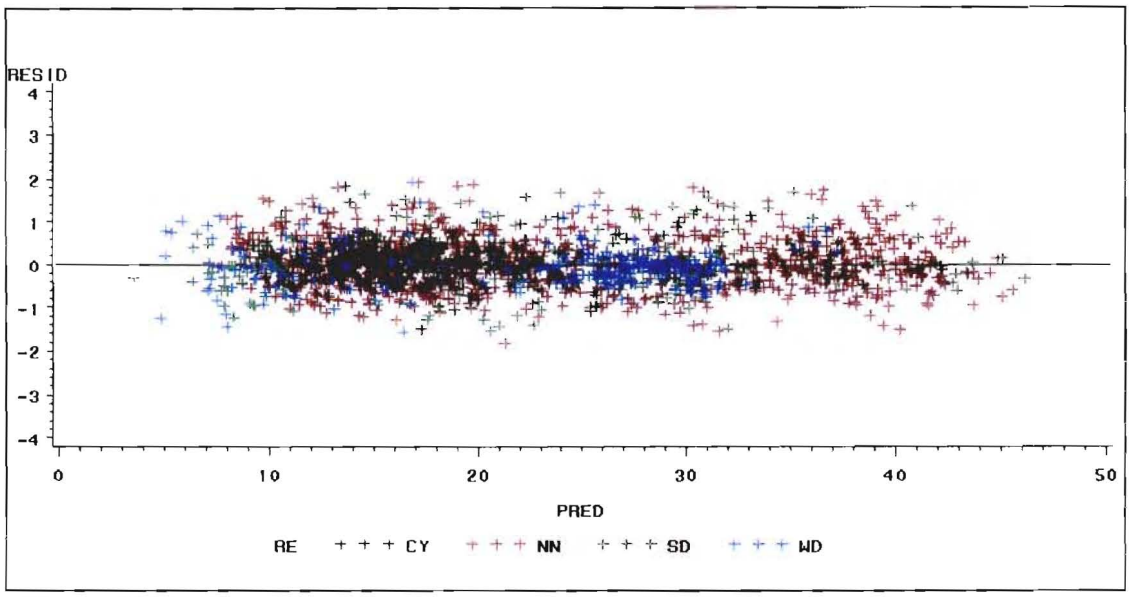

Figure 7.10: Plot of residuals vs predicted value for the short interval mean top height equation with new model

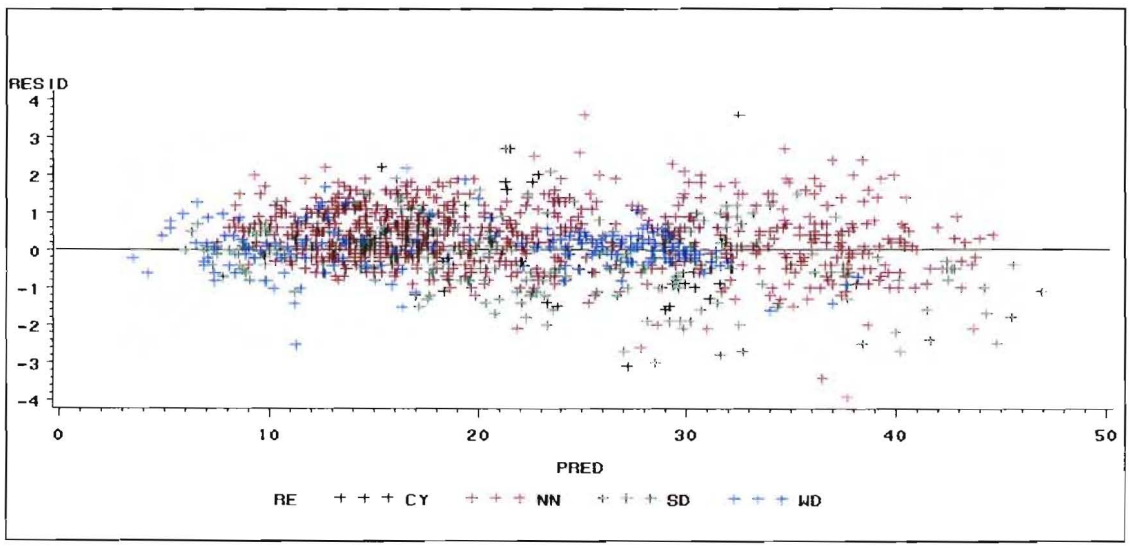

Figure 7.11: Plot of residuals vs predicted values for the short interval mean top height equation with SIDFIR

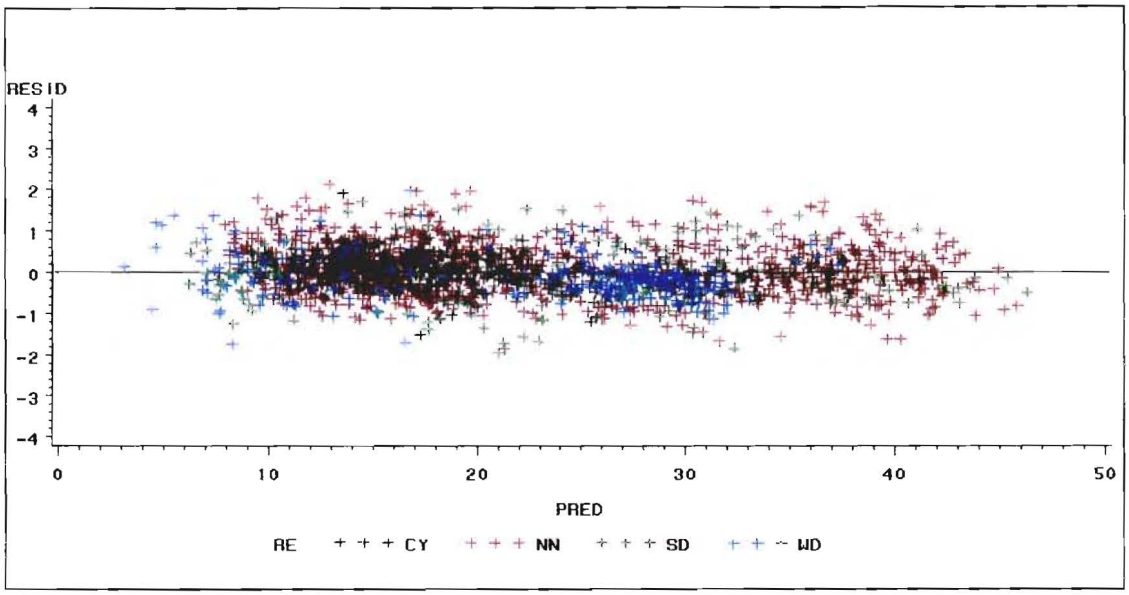

Figure 7.12: Plot of residuals vs predicted values for the short interval mean top height equation with DirfirStand

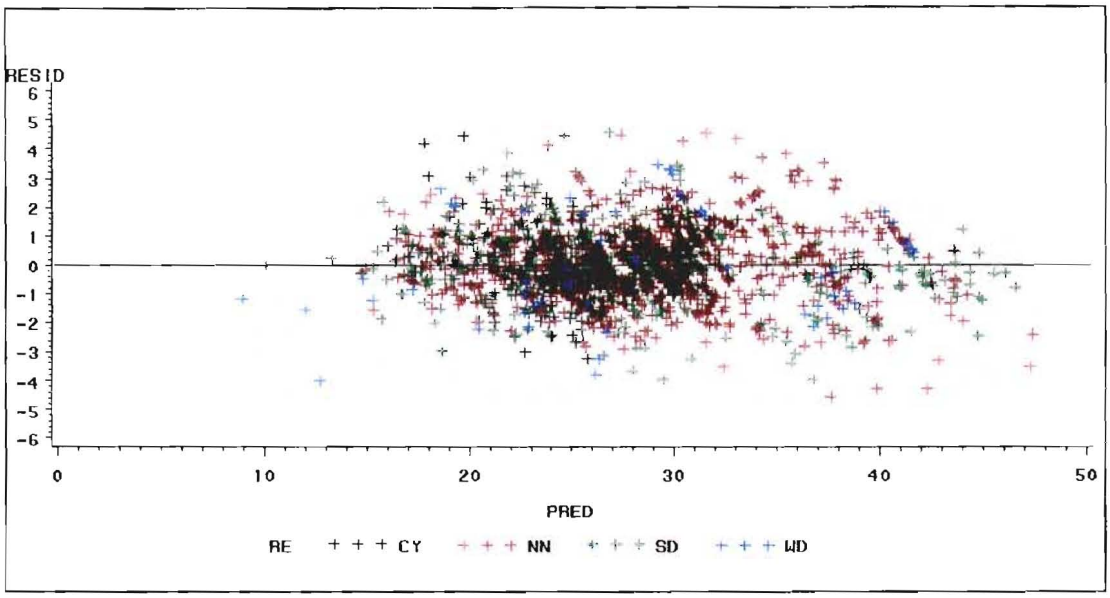

Figure 7.13: Plot of residuals vs predicted values for the long interval mean top height equation with new model

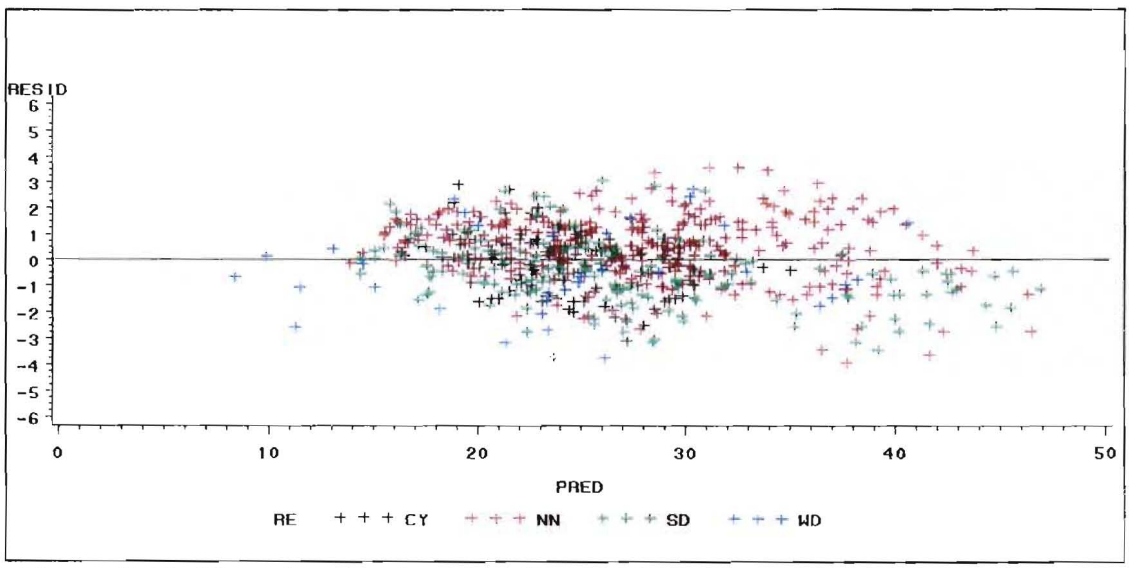

Figure 7.14: Plot of residuals vs predicted values for the long interval mean top height equation with SIDFIR

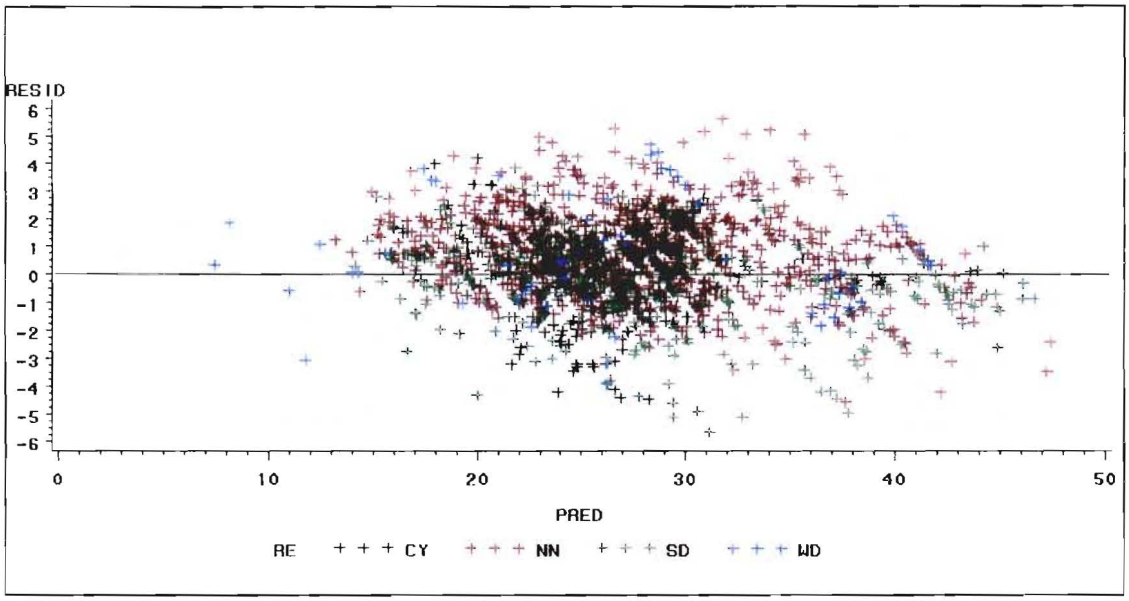

Figure 7.15: Plot of residuals vs predicted values for the long interval mean top height equation with DirfirStand

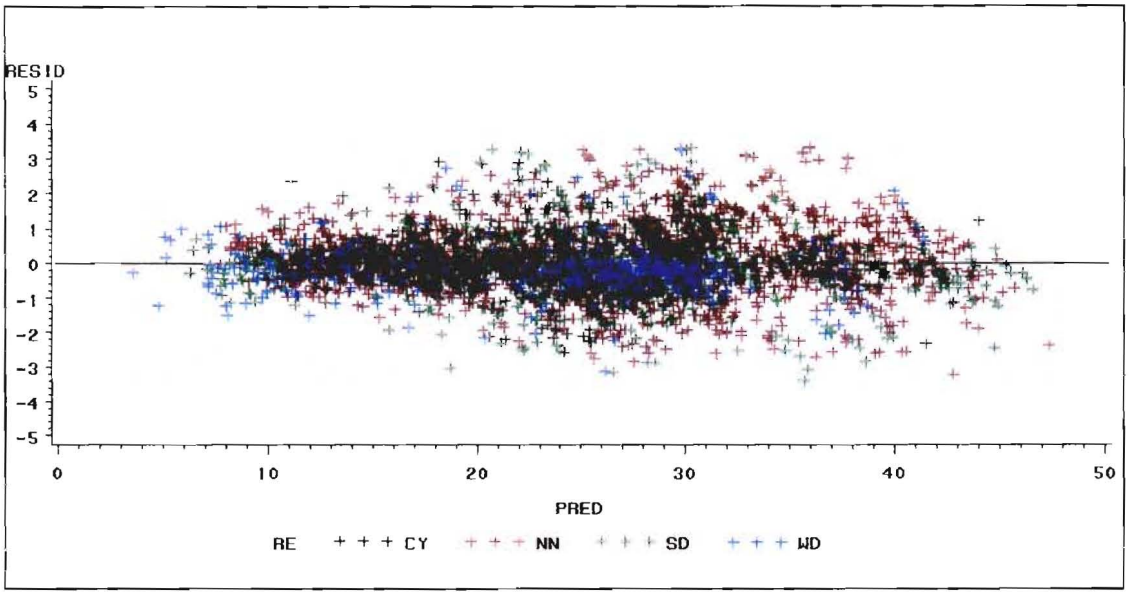

Figure 7.16: Plot of residuals vs predicted values for the mixed interval mean top height equation with new model

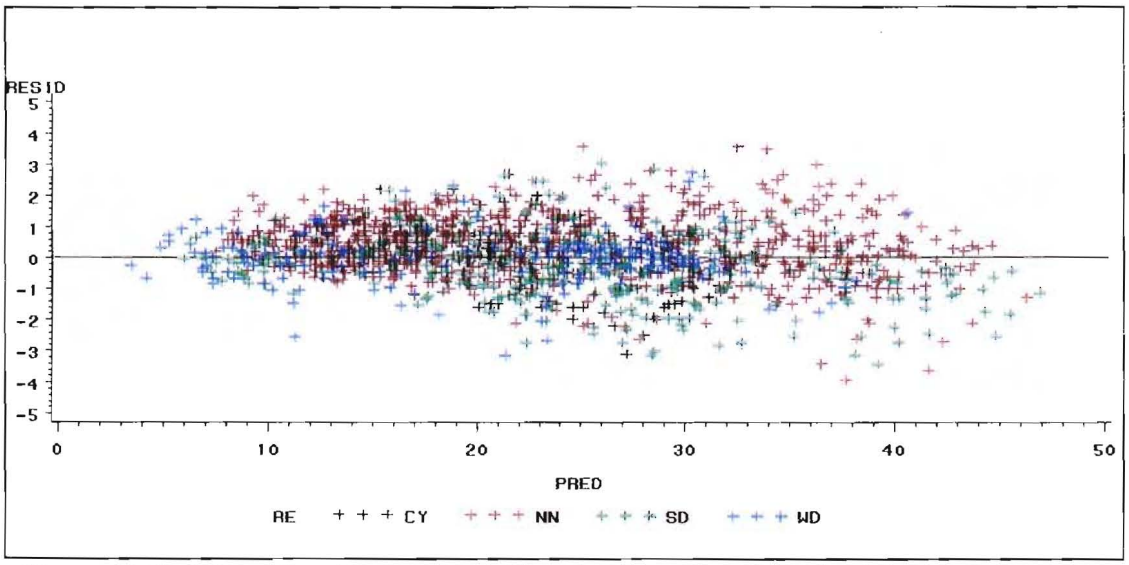

Figure 7.17: Plot of residuals vs predicted values for the mixed interval mean top height equation with SIDFIR

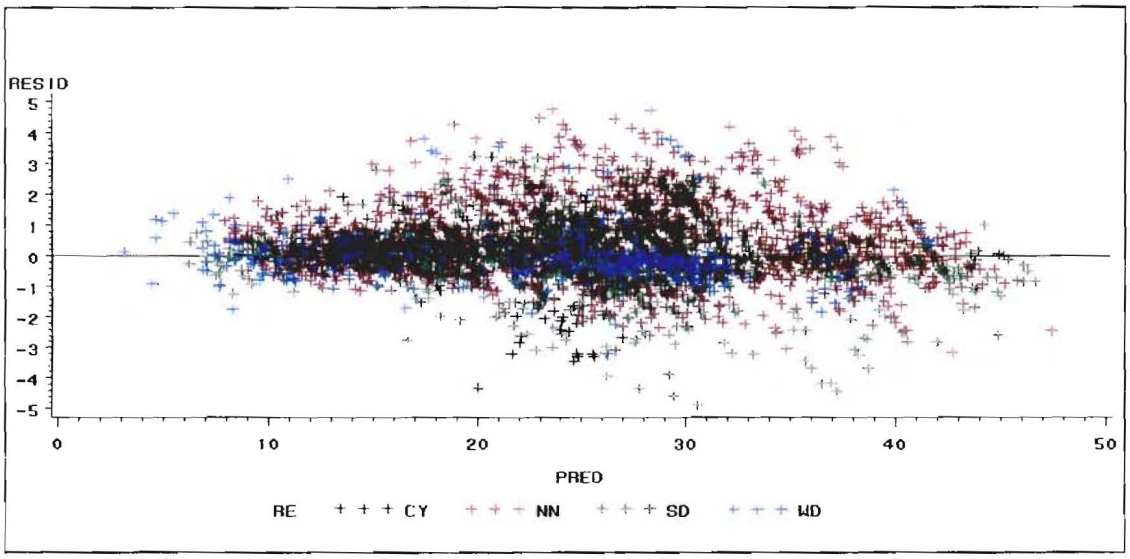

Figure 7.18: Plot of residuals vs predicted values for the mixed interval mean top height equation with DirfirStand

(3) Survival of Stems/ha Equation

The short interval model had mean residual -0.674 stems/ha, implying slight over-prediction with an average deviation 28 of stems/ha. The corresponding value for DirfirStand was –2.314 with a deviation of 30 stems/ha and for SDFIR was 9.76 with deviation of 91 stems/ha. SIDFIR predicted more stems/ha than DfirStand and the new model. The three models had similar mean residual errors for the long interval data, but SIDFIR showed less precision with extreme residuals –592 and 512 stems/ha than New and DirfirStand.

The mean residual of the mixed interval mortality equation was 1.63 stems/ha, indicating slight under-prediction with an average deviation of 72.9 stems/ha. These estimates were more precise than those of the SIDFIR and DirfirStand models, which had mean residuals of –1.57 stems and 0.81, and with deviations of 107.3 and 76.7 stems/ha respectively. Figures from 7.19 to 7.27 show the residuals patterns for the models.

Comparisons of the three models showed that the new model showed considerable overall improvement in the statistics over SIDFIR and DirfirStand for all equations in each interval data set.

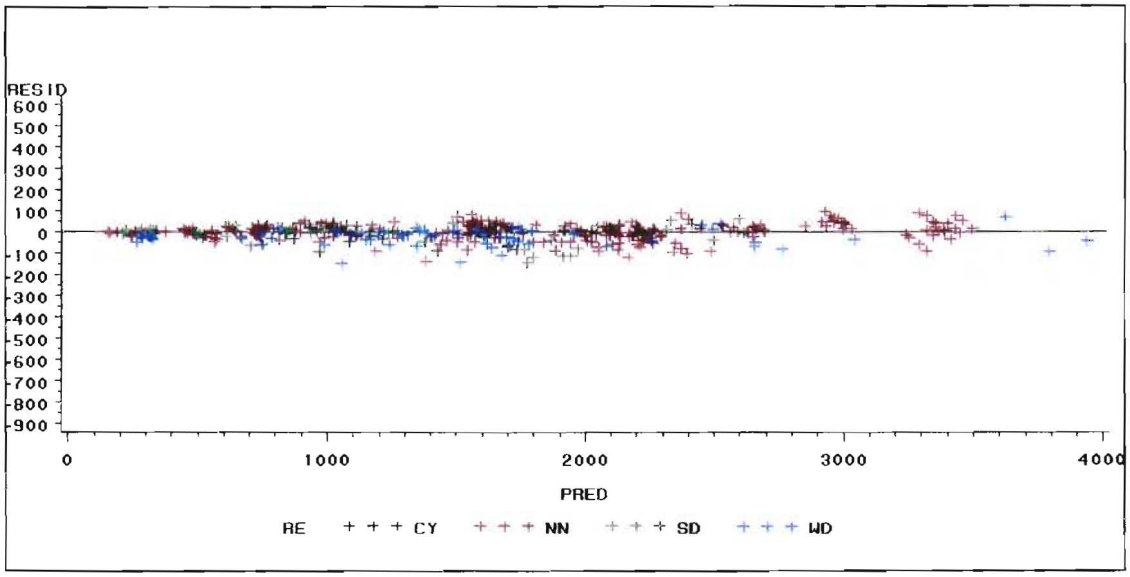

Figure 7.19: Plot of residuals vs predicted values for the short interval stem survival/ha equation with new model

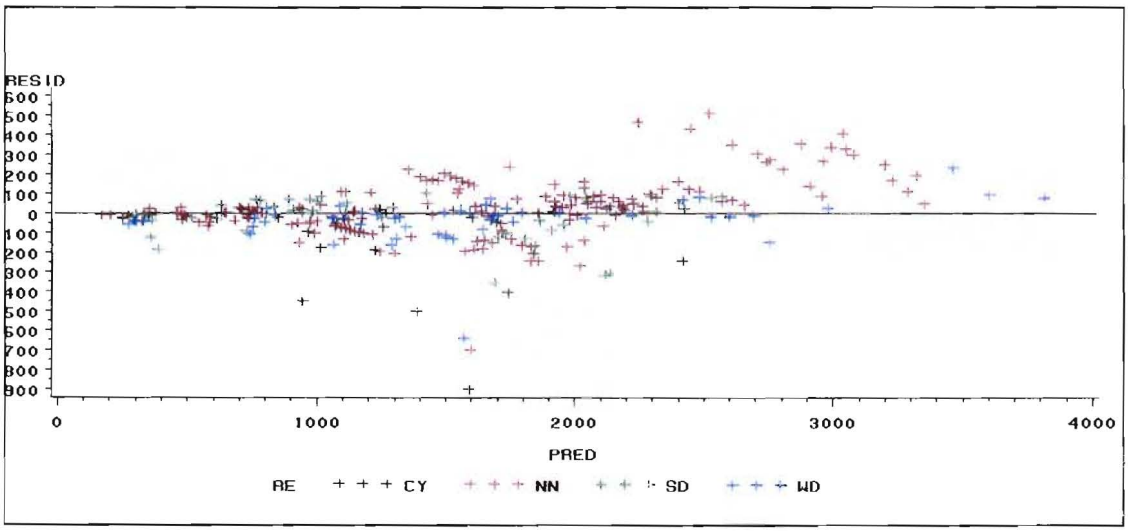

Figure 7.20: Plot of residuals vs predicted values for the short interval stem survival/ha equation with SIDFIR

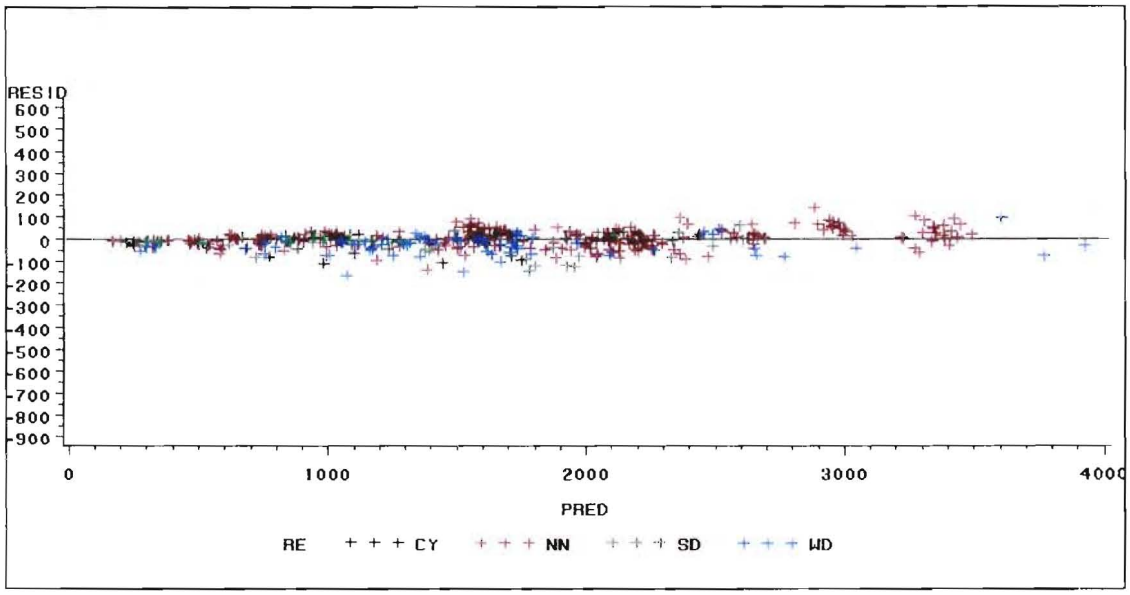

Figure 7.21: Plot of residuals vs predicted values for the short interval stem survival/ha equation with DfirStand

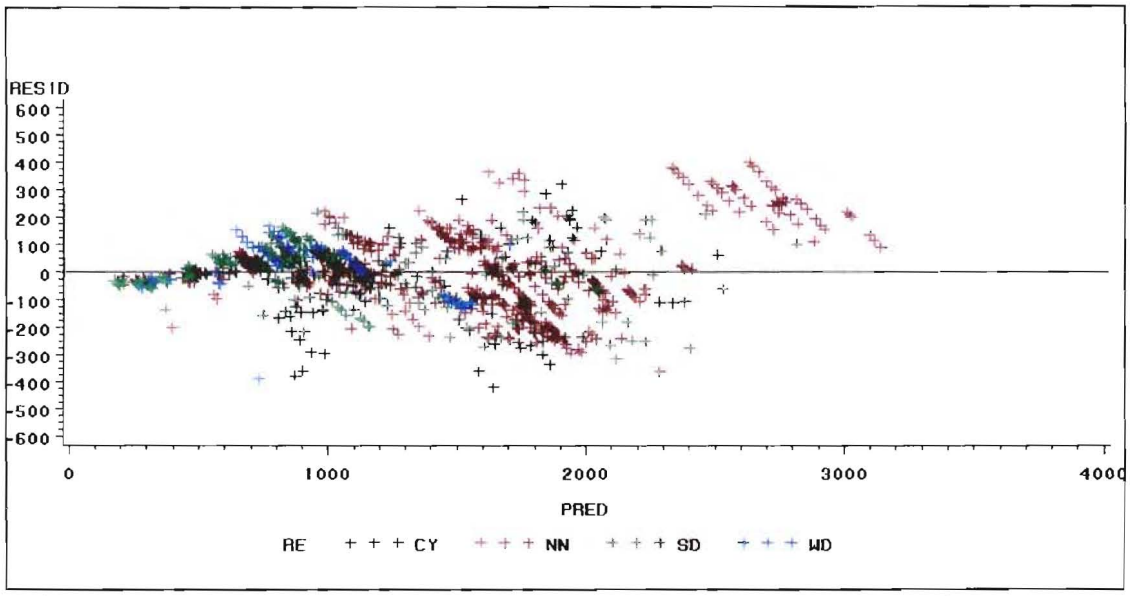

Figure 7.22: Plot of residuals vs predicted values for the long interval stem survival/ha equation with new model

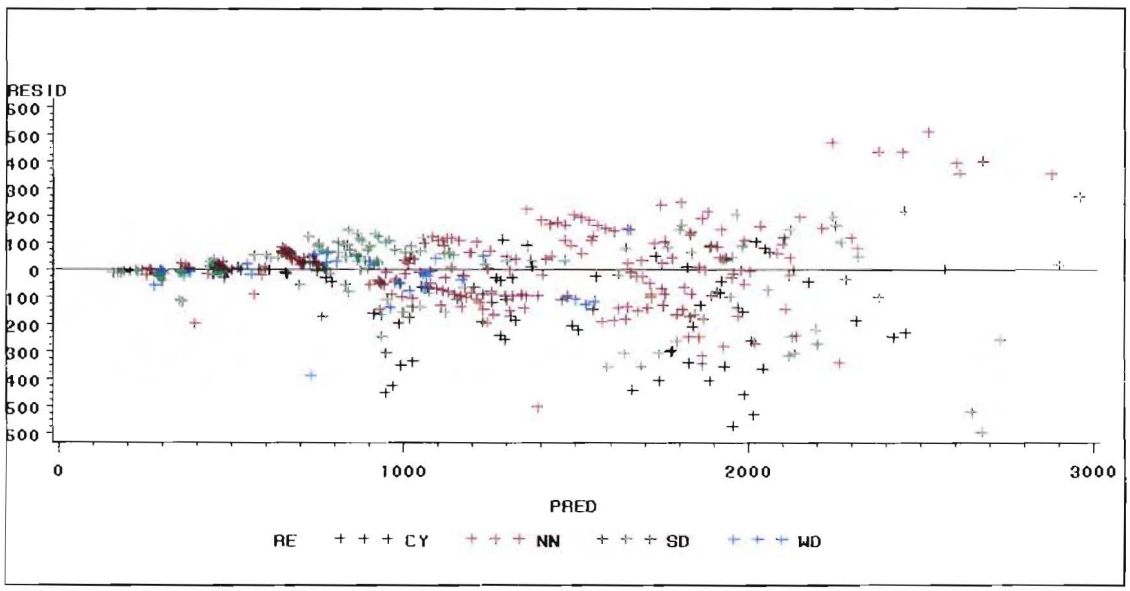

Figure 7.23: Plot of residuals vs predicted values for the long interval stem survival/ha equation with SIDFIR

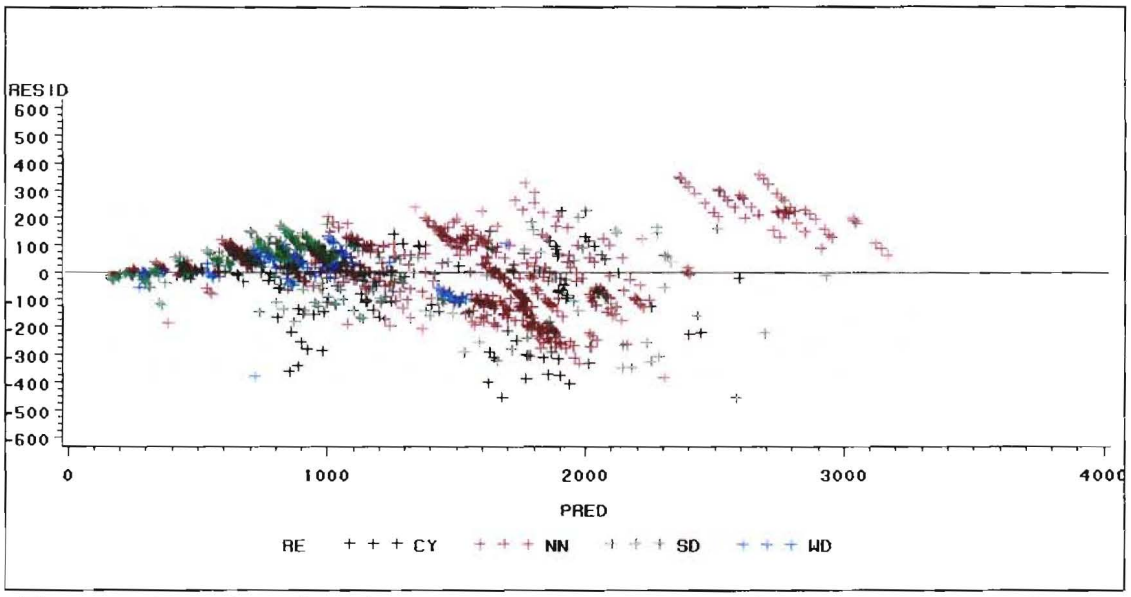

Figure 7.24: Plot of residuals vs predicted values for the long interval stem survival/ha equation with DirfirStand

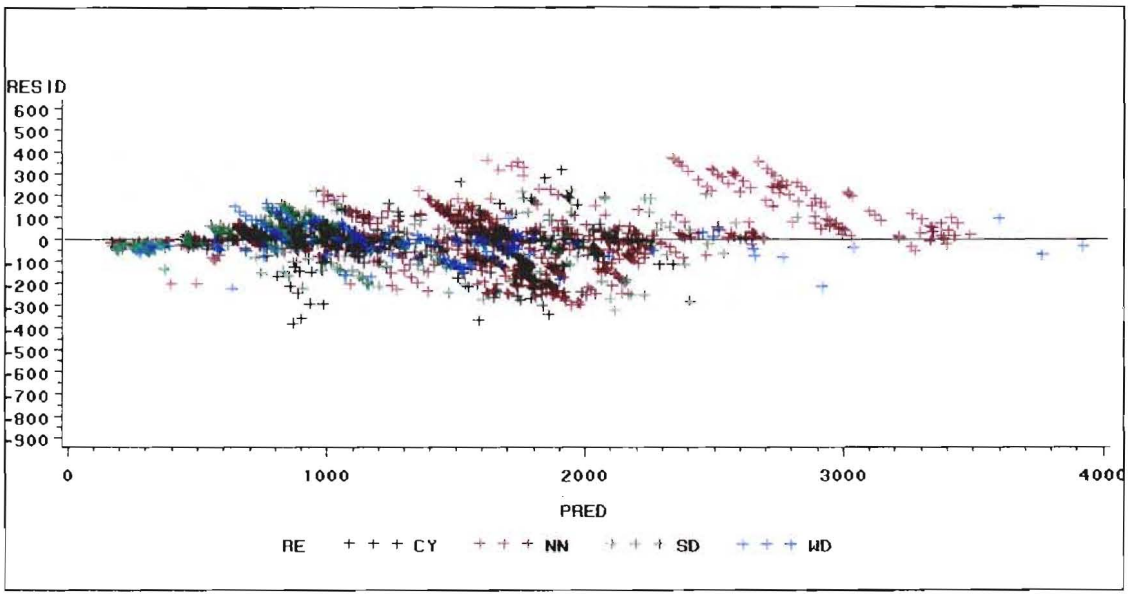

Figure 7.25: Plot of residuals vs predicted values for the mixed interval stem survival/ha equation with new model

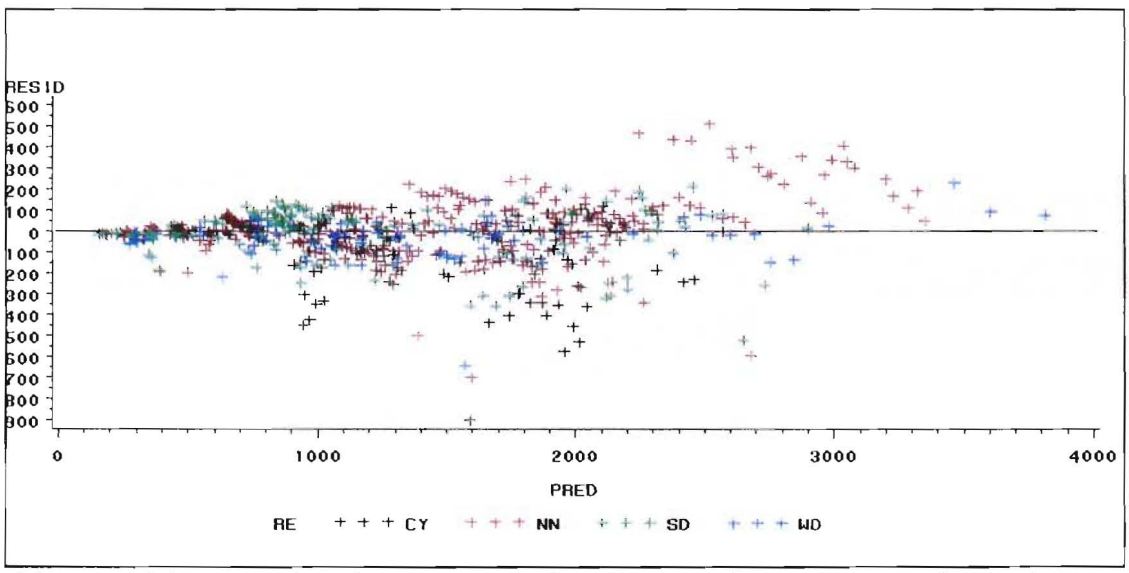

Figure 7.26: Plot of residuals vs predicted values for the mixed interval stem survival/ha equation with SIDFIR

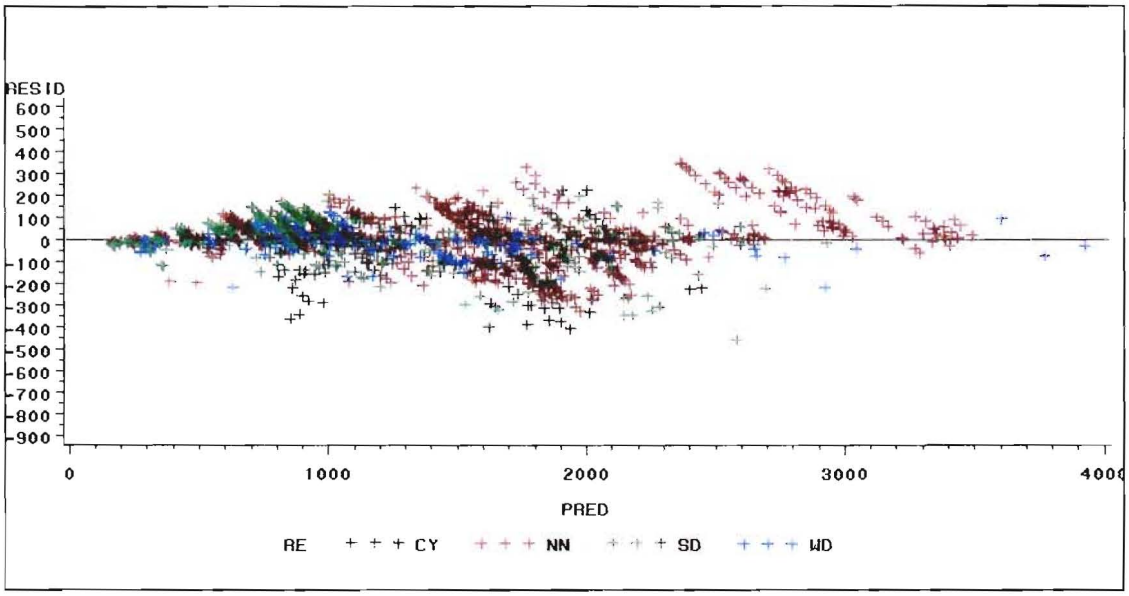

Figure 7.27: Plot of residuals vs predicted values for the mixed interval stem survival/ha equation with DirfirStand

Growth and yield prediction by the best new model, DfirStand and SIDFIR had been compared in order to assess the reliability of the new models. The simulations were done with initial starting conditions; the initial measurements of 20 years, site index of 30 m, basal area of 30 m^2/ha, 1000 stems/ha and a mean regional altitude. These initial measurements were used for each region, projected and compared at age 45 years with the results of DfirStand and SIDFIR. Table 7.2 summarises simulation results for these three models.

Table 7.2 Comparison result of simulation at age 45 years for new model, DfirStand and SIDFIR

Model	Basal area/ha (m^2/ha)	Mean top height (m)	Stocking (stems/ha)
New Model (CY)	70.4	32.8	641
New Model (NN)	82.1	33.3	781
New model (SD)	89.4	33.1	744
New Model (WD)	90.4	33.3	755
DfirStand (CY)	76.3	33.1	728
DfirStand (NN)	77.5	33.0	728
DfirStand (SD)	91.0	33.2	728
DfirStand (WD)	92.8	33.1	728
SIDFIR	80.8	33.3	784

where, CY, NN, SD and WD represent Canterbury, Nelson Southland and Westland respectively.

The new model predicts almost same mean top height for site index 30 m as DfirStand and SIDFIR at age 45 years. The prediction of basal area/ha for SIDFIR was accomplished using average values. The new model and DfirStand predict basal area/ha separately for different four regions using of dummy variables representing regions, and regional altitudes. Thus, Westland has the highest production of basal area/ha, followed by Southland, then Nelson and Canterbury the least in both the new model and DfirStand. The prediction of stem survival/ha for the new model varied from region to region through dummy variables, but the other two models predict

stem survival/ha of average values for the whole South Island. The predicted stem survival/ha for the Canterbury region is lower than the other regions in the new model, but the other regional predictions for the new model are similar to those of the other two models. The three models differ in basal area/ha, mean top height and stem survival/ha predictions through localization by dummy variables.

7.2 Logical Properties of Projection Equations

7.2.1 Net Basal Area/ha Equation

The net basal area equation used for the mixed interval data set was a modified form of Gompertz function.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))}$$

The predictions of basal area/ha vary among the four regions. Different growth patterns were represented using dummy variables which affected the levels of asymptotes. Altitude was also found to represent variability in site quality. Explanatory variables were carefully introduced in ways that ensured the desirable logical properties of equations were not compromised. The consistency property is that when $T_1=T_2$ then $G_1=G_2$ and as T_2 tend to ∞, G_2 approaches an upper asymptote $(\alpha+\beta Ki)$ for any specific region. The equation had path invariance whereby the projection from T_1 to T_3 gave the same results as two step projection s from T_1 to T_2 followed by T_2 to T_3.

7.2.2 Mean Top Height Equation

The mean top height equation for the mixed interval data set was represented by a Schumacher polymorphic form.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/T_2)^{\beta})$$

The asymptotes of this equation changed from α to $(\alpha+K\beta i)$ according to each of the four different regions. The equation is logically consistent in that as $T_1=T_2$ then $h_{100,1}=h_{100,2}$ and as T_2 approaches ∞ then $h_{100,2}$ approaches upper asymptotes. A single projection from $h_{100,1}$ to $h_{100,3}$ will give the same results as a projection from $h_{100,1}$ to $h_{100,2}$ and then to $h_{100,3}$.

7.2.3 Stem Survival/ha Equation

A modified Gompertz function of the form presented here was found to fit survival stems/ha best for the mixed interval data.

$$N_2 = N_1\, e^{(\beta(T_2^2-T_1^2))}\, e^{(\alpha+\beta_1K_1+\beta_2K_2+\beta_3K_3+\beta_4S)(1-\exp(\beta(T_2^2-T_1^2)))}$$

The asymptotic value of this equation varied with site index and dummy variables, meaning that survival was different for each region, and stands on sites with higher site indices had more mortality than those on sites with lower ones. The equation contained all desirable properties in that as $N_1=N_2$ when $T_1=T_2$ and the projection from T_1 to T_3 is the same as the projection from T_1 to T_2 and then to T_2 to T_3.

In order to demonstrate that the projection was biologically realistic, simulations were made using the best mixed interval based equations with given initial starting conditions (e.g. site index of 30 m, basal area of 35 m^2/ha, mean top height of 16.3 m, 1000 stems/ha, the initial measurements of 20 years and a mean regional altitude). Table 7.3 represents the results of simulation for the new model. Basal area/ha, mean top height predictions increased with age and stem survival/ha predictions decreased as time elapsed.

Table 7.3 Summary of output from simulation for best new model

Region (Mean altitude)	Age (years)	Basal area/ha (m^2/ha)	Mean top height (m)	Stocking (stems/ha)
Canterbury (330 m)	20	35.0	16.3	1000
	25	46.4	20.2	932
	30	55.9	23.8	859
	35	63.3	27.1	784
	40	69.0	30.1	710
	45	73.2	33.0	641
	50	76.4	35.6	576
	55	78.7	38.1	518
Nelson (441 m)	20	35.0	16.3	1000
	25	49.2	20.8	962
	30	61.5	25.0	919
	35	71.6	28.9	874
	40	79.4	32.6	827
	45	85.3	36.1	781
	50	89.8	39.4	737
	55	93.0	42.5	695
Southland (217 m)	20	35.0	16.3	1000
	25	50.8	20.5	954
	30	65.0	24.4	904
	35	76.6	28.0	851
	40	85.8	31.4	797
	45	92.9	34.6	744
	50	98.2	37.5	693
	55	102.1	40.3	645
Westland (228 m)	20	35.0	16.3	1000
	25	51.0	20.9	957
	30	65.4	25.2	909
	35	77.4	29.2	858
	40	86.8	33.0	806
	45	94.0	36.6	755
	50	99.5	39.9	706
	55	103.5	43.1	660

7.3 Limitations of Application of the Models

The range in age of stands is from 5 to 78 years. The models should be applied to stands of these ages. Outside these ranges, users of the model require extreme caution. Differences in net basal area/ha, mean top height and survival stems/ha growth patterns across regions are explicitly represented for each equation. The

models should apply to forests grown in the South Island. The models provide reliable results for forests grown in Canterbury, Nelson, Southland and Westland.

Chapter 8

Discussion

This study has produced equations for a stand growth and yield model for plantation crops of Douglas-fir grown throughout the South Island of New Zealand. A computer program for this is appended. The methodology for developing equations in this study allows for re-arrangement of the data to different interval lengths between measurements. This way of modelling stand growth and yield was tried to see if sensitivity gains could be made and to identify whether sound models could be produced using only short intervals or alternatively only long ones.

The new model developed here is a whole stand growth and yield model which involves fitting single equations to compatible projections of mean top height (h_{100}) and net basal area/ha and stem survival/ha (N) for Douglas-fir stands grown in Canterbury, Nelson, Southland and Westland regions separately.

All selected equations represented a precision which is at least at the same level of achievement as, but mostly better, than existing models.

8.1 Features of the Study that are New

8.1.1 Use of Different Interval Lengths

Before development of this new model for Douglas-fir grown in the South Island of New Zealand, the convention used in developing growth and yield models was mostly to use successive re-measurements, which mostly measured short projection intervals, although Villanueva (1992) used all possible interval lengths data for Caribbean pine in Fiji without evaluating the benefits of it. The alternative strategy approach adopted here, evaluated the use of three different interval data sets; short, long and mixed intervals, which would provide a basis for selecting an appropriate combination of projection intervals to improve the reliability of growth and yield predictions for Douglas-fir in the South Island and assist model developers in future.

This approach has shown that the precision of estimated coefficients could be increased by using a data set which contains a sufficiently representative range of projection intervals. In comparing coefficients of the short with the mixed interval data based-equations, the latter presented a much tighter 95 percent confidence interval. When these two equations were applied to the long interval data set, the trend of residual pattern of the mixed interval model was tighter than that of short interval equation. Other statistics such as mean residual error and skewness, obtained from PROC UNIVARIATE procedure, showed that the mixed interval equation had lower values. Thus, gains in narrowing the distribution of residuals by using different interval lengths are considerable for growth and yield studies, and offer potential ability to expand data sets when the successive interval data are too few for traditional growth and yield study.

The application of the best short interval equations to the long and mixed interval data revealed biased residual patterns for net basal area/ha and stem survival/ ha equations,

while application of the best mixed interval models to the other data sets showed good fits without apparent bias. The cross-fitting for mean top height showed consistency among the predicting equations, because the best equation for different interval data sets had same functional form and the only differences was that two additional dummy variables representing locality were needed for the long and mixed interval data sets.

Cross fitting showed that the equations from the mixed interval data set, which included interval lengths from short to long, fitted well for all data sets. Using only shorter intervals lead to excessive variance in regression functions. For example, in a Schumacher projection equation basal area/ha (G_2) will be projected at time T_2 following way.

$$\ln (G_2) = \ln(G_1) (T_1/T_2)^{\beta} + \alpha(1-(T_1/T_2)^{\beta}) \quad (8.1)$$

Equation 8.1 can be written as equation 8.2

$$\ln (G_2) - \ln (G_1) (T_1/T_2)^{\beta} = \alpha(1-(T_1/T_2)^{\beta}) \quad (8.2)$$

Equation 8.1 and 8.2 are form of $Y = \alpha X^{\beta}$ which is $Y = \alpha X$ when $\beta = 1$.
The variance of α is given by

$$V(\alpha) = \sigma^2 / (\sum (X_i - \mu)^2) \quad (8.3)$$

where,
μ = the arithmetic mean
$(X_i - \mu)$ = the deviation of an individual measurement from the mean of all measurement.

Where the mean proportional time (μ) and the measurement time are close (i.e. short interval), the variance will be large, while the difference large, the variance will be smaller, and then precision of estimate greater. Considering $Y = \alpha X^{\beta}$, where X = (1-

(T_1/T_2)), when T_1 and T_2 are closed this give a restricted range of X, for example, $T_1 = 4$ and $T_2 = 5$ then $X = \alpha(1-(4/5) = 0.2\alpha$. While various ranges of T_1 and T_2 from the closest to widest provide a large range of X, for example, when $T_1 = 4$, $T_2 = 5$, $T_3 = 6$, ..., and $T_{n+1} = 25$, the closest is $T_1 = 4$ and $T_2 = 5$ then $X = \alpha(1-(4/5) = 0.2\alpha$, the widest is $T_1 = 4$ and $T_2 = 25$ then $X = \alpha(1-(4/25) = 0.84\alpha$.

Forecasting forest growth and yield is the process of conjecture of the future based on the past and present data, which requires history of the stands through time lapse. With short interval measurement, increasing measurement intensity might not result in coefficient estimates that are more precise. Thus, appropriately duplicated time interval data may represent better characteristics of the stand than only short interval data. This result shows that developing projection equations using mixed interval data sets was sufficient not only to obtain a precise model but also a model applicable to other data sets.

Permanent sample plots are commonly re-measured at intervals of 1 to 10 years (Vanclay, 1994; Avery and Burkhart, 1994; Husch, 1971), depending on timber growth rates, expected changes in stand conditions, and the intensity of management. The interval should be long enough to permit a measurable degree of change, but short enough so that a fair proportion of the trees originally measured will be present for re-measurement. Conversely, cost efficiencies demand that re-measurement should not be unnecessarily frequent. Thus, the implications of using mixed interval data sets allows the utilization of maximum growth information from permanent sample plots and minimizes the cost of measuring operations.

8.1.2 Use of Gompertz Function

A Gompertz function was frequently found to be the best growth and yield projection equation in this study. There is no particular sigmoid equation which is best for all situations. Woollons et al. (1990) pointed out that it is unrealistic to expect a unique

function to perform consistently better than others with forest growth and yield data. The selection of appropriate equations and exploring their utility are important for the goodness of fit of models, with nature of the data varying with individual circumstance. It is likely that the applicability of the sigmoid function depends on the data set being used. The Gompertz function represents T_1 and T_2 as T_2-T_1, while the Schumacher function, which was found to be suitable for the short interval data set, encapsulates T_1 and T_2 as T_2/T_1. The applicability of the Gompertz function, therefore, is enhanced with data sets that contain longer projection intervals. If short interval data sets were employed for the Gompertz function, term T_2-T_1 would take on values of only one and two, and then have a reduced effect on precision of the equation and an adverse effect on estimating coefficients and projecting long term yields. When short interval data sets were used for fitting models the Schumacher function appeared to be more suitable because T_2/T_1 varies more widely than the difference between T_1 and T_2 for the Gompertz function.

8.1.3 Use of Dummy Variables

The existing NZFRI model for Douglas-fir grown in the South Island provides average predictions of yield for the entire South Island region, while Temu's model offers different yields by region only for net basal area/ha production. In this study, however, dummy variables representing different regions were found useful for each of the three state variables (net basal area/ha (G), mean top height (h_{100}) and stem survival/ha (N)).

The dummy variable approach allows one to use a single equation for all regions, provided that there is comparability in the growth trajectories and simply a difference in asymptotes. For example, when modelling net basal area/ha predictions with equation 8.4, the overall equation for basal area/ha is disaggregated to provide different projection equations for each of the four regions individually.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (8.4)$$

Equation 8.4 is applied as default and equation 8.5 is for projections for the Canterbury region.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (8.5)$$

Equation 8.6 includes an active dummy coefficient β_1 and variable K1 for the Nelson region.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_1 K_1)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (8.6)$$

Similarly, equation 8.7 contains active $\beta 2$ and K2 for the Southland region.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_2 K_2)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (8.7)$$

Lastly, equation 8.8 represents active $\beta 3$ and K3 for the Westland region.

$$G_2 = G_1 e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} e^{(\alpha+\beta_4 ALT+\beta_3 K_3)(1-\exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (8.8)$$

There has been a tendency in many countries to prescribe growth and yield models even for very small populations (Whyte et al., 1992). This approach which results in a proliferation of models is frequently not justified. Preliminary work in fitting equation

for sub-populations showed that coefficients for the sub-population did not vary significantly to those of aggregated populations.

For example, when equation 8.4 without dummy variables was fitted to each region, coefficients for each region were very similar to coefficients for the whole region, as shown in Table 8.1. Consequently, using dummy variables identified that there were different growth trends for net basal area/ha, mean top height and stem survival/ha in Canterbury, Nelson, Southland and Westland. The gains through using these dummy variables are that regional growth trends are identified and these provide forest managers with the opportunity to make more region specific management decisions based on a much larger data set.

Table 8.1 Comparison coefficients for each region and for whole region

Region	α	β	γ
Canterbury	4.72	0.10	0.001
Nelson	4.79	0.09	0.001
Southland	4.71	0.11	0.007
Westland	4.98	0.09	0.007
Whole region	4.78	0.09	0.006

Comparisons of yields per hectare predicted regionally with some starting conditions are described in Chapter 7 and are discussed further here. The differences in mean top height and net basal area regionally simulated are expressed in terms of volume/ha at the rotation age with prediction from the Law's model as denominator. The stand volume equation used in this study was;

$$V = 0.5909\, G^{0.9795}\, h_{100}^{0.8624} \qquad (8.9)$$

which was developed by Temu (1992) and shown to be very precise. Three different stand conditions were used as initial variables:

1) initial measurement of 20 years, site index of 30 m, basal area of 30 m^2/ha, 1000 stems/ha and a mean regional altitude;

2) initial measurement of 15 years, site index of 22.5 m, basal area of 29.5 m^2/ha, 2300 stems/ha and a mean regional altitude;

3) initial measurement of 27 years, site index of 25.1 m, basal area of 56.45 m^2/ha, 2100 stems/ha and a mean regional altitude.

The effects of localization of the new model for height prediction were not so great as shown Table 8.1. It forecasts height in almost the same way for the four different regions, and is similar to that of SIDFIR. The differences in predictions for basal area/ha by region however, were large especially for the Canterbury region that had lower production of basal area compared to the other regions. A comparison of basal area/ha projections of the new model with SIDFIR was made. The former predicted higher values for Nelson, Southland and Westland at rotation ages of 40 years in the first and second stand conditions, and for Southland and Westland in the third stand condition. In stem survival/ha prediction, the Canterbury region was found to suffer more mortality than the other regions. Westland had the highest production of stand volume/ha, followed by Southland, then Nelson, and Canterbury at a rotation age of 40 years. The new model predicted more volume per hectare for Southland and Westland by 5.5 % and 3.4%, 6.4% and 4.2% than SIDFIR at a rotation age of 40 years in the first and second stand conditions. While the latter predicted 15 %, 19% and 10% more volume per hectare for the Canterbury region in the first, second, and third stand conditions respectively. A summary of these comparisons is presented in Table 8.2.

Table 8.2 Estimates of yield comparison of new model with SIDFIR

Stand Condition	Region	Age (years)	Basal area (m^2/ha)	Height (m)	Stocking (stems/ha)	Volume (m^3)	Percent (%)
1	CY	40	65.7	30.0	710	669.6	-15
	NN	40	75.7	30.0	827	767.9	-2.6
	SD	40	81.8	30.1	797	831.4	5.5
	WD	40	82.7	30.0	806	838.2	6.4
	SIDFIR	40	72.6	30.0	828	788.0	0
2	CY	40	73.1	22.5	1465	580.0	-19.1
	NN	40	85.7	22.6	1739	680.1	-5.1
	SD	40	93.7	22.5	1666	741.1	3.4
	WD	40	94.9	22.4	1688	747.4	4.2
	SIDFIR	40	85.5	22.5	1757	717.0	0
3	CY	40	73.0	25.2	1533	637.2	-10.1
	NN	40	80.9	25.1	1718	703.1	-6.1
	SD	40	85.6	25.2	1670	745.6	-0.5
	WD	40	86.3	25.1	1685	750.5	0.2
	SIDFIR	40	81.0	25.1	1787	749.0	0

Site quality, stocking variability, genotype variability and local climatic fluctuations among different regions, can affect growth and performance of individual trees and forests at the stand level, resulting in different growth trends between regions. This study confirmed these trends.

8.2 Growth and Yield Models with a Data Set Free from Auto-Correlation

When data are obtained from repeated measurement on the same individual sample plots over time, as is the case with PSP data, there are correlated errors. This will result in underestimation of the standard error of least-squares regression coefficients. In practice, however, the least-squares method has been considered to be appropriate for estimation coefficients, provided that statistical analysis and interpretation are more considered to estimating residuals to check bias in fitting, and there is assurance that logical relationships are reflected in fitting equations and sensible criteria for including independent variables are adopted. Additionally, a data set which has much less correlated data can be used to examine whether equations so derived are similar to those from the corresponding ones derived from full data sets.

In this study uncorrelated data were obtained by selecting one re-measurement from PSPs which were then used to check consistency of coefficients between data set free from auto-correlation and full data sets. When equations which were found to best fit the data set free from auto-correlation were fitted to the full data set it was evident that they had very similar regression coefficients in basal area/ha, mean top height and stem survival/ha equations. The coefficients were significantly different by each region in data set free from auto-correlation and full data sets, and the coefficients were significantly different from zero. In statistical point of view, it would be appropriate to require no correlation among the data, but in practice this may be of little importance provided that certain circumspection.

8.3 Possible Further Refinements to the Model

1) The equations for Canterbury and Westland region were based on somewhat limited numbers of observations: Canterbury was represented by 10% and Westland was

represented by 13% of the all observations. More data are needed in the future from these regions to provide more reliable forecasts.

2) No account of the presence of the Swiss needle cast fungus (Phaeocytopus gaeumannii) has been taken in this study, though it is known to spread to Douglas-fir grown throughout the South Island region. Future work may be needed to address and evaluate the import of this disease, along the lines outlined by Liu Xu (1990) for Central North Island Douglas-fir. In that study, Liu Xu fitted separate basal area/ha equations for diseased and undiseased stands. The basis of separating data to disease and undiseased stands was the time that the infection occurred (e.g. pre-1963 and post-1963). However, there were no exact records about the disease in the PSP data set, so basic time of infection derived from previous studies (Beekhuis, 1978; James and Bunn, 1978), which showed that there was no obvious growth loss data because of the disease prior to 1963. If information about the exact time of disease infection was available for the South Island, a more refined prediction for Douglas-fir growth could be forecasted.

3) Dummy variables were used mainly to differentiate the geographic locations of the different four regions. Other aspects which influence tree growth such as, soil aspect and climatic factors within each region could be used to refine predictions in future studies.

4) Further work is needed for improving the goodness of fit for the stem survival/ha equations. Data which are well balanced regionally need to be obtained.

8.4 Recommendations

8.4.1 Applicability of Equations

The equations apply to even-aged Douglas-fir stands aged from 5 to 80 years grown in four South Island regions. Users who wish to use the overall model to make predictions for Douglas-fir plantations should utilise regional crop production data and monitor predicted forecasts. The initial inputs of mean top height can be used for estimating site index in the model. All input values are required with one decimal palace, except initial values of stems/ha which are integers. Input values required are similar to most existing growth and yield models in New Zealand.

8.4.2 Data Format

The use of a mixed data set, which included an adequate a range of measurement periods has been found to increase precision in estimating model coefficients without bias in this study. Hence, this kind of data format would be suitable for further growth and yield studies. In order to create mixed interval data format refer to the program (Figure 3.3) which was used in this study and make sure not to include of combinations of re-measurements which include thinning within the projection interval.

In New Zealand it is common practice to measure plots annually or biannually. When permanent sample plots have been re-measured with short intervals, it is desirable that the data be transformed to allow various measurement intervals to be incorporated to increase precision in estimating model coefficients.

Difference equation methods with different interval length data were found to provide better predictions than those of Garcia's state-space approach (1988, 1994) which predict state variables simultaneously as, for example, in the SIDFIR model. Gains in

narrowing the distribution of residuals and more precise coefficient estimation by using different interval lengths are considerable for growth and yield studies.

Chapter 9

Summary and Conclusions

This study investigated ways and means of refining an existing stand growth model for Douglas-fir even-aged stands distributed throughout the South Island of New Zealand which was developed by Temu (1992). Analyses showed that, by forming three data sets which represented different interval lengths between the re-measured growth data, it was possible to derive much more precise and robust models than those achieved using successive measurement intervals. The models developed here contain several features that appear to have improved precision and reliability of estimates of future yield.

9.1 Whole Stand Growth Model

The data representing periodic re-measurements spanning 1 to 28 years were separated into the following three sets, which reflected different interval lengths between re-measured variables:

(a) short interval, 1 to 3 years between measurements;
(b) long interval, 8 to 28 years between measurements;
(c) mixed above long and short intervals.

Three stand variables were examined, namely basal area/ha (G), mean top height (h_{100}) and stem survival/ha (N). In almost all previous studies of growth and yield, short interval data sets with consequent narrow interval periods have been employed. In this study, re-arrangement of the data created an adequately replicated range of years between re-measurements to be investigated in order to evaluate which ones could provide more precise coefficient estimates of growth and greater accuracy of yield predictions.

In comparing coefficients of the short with the mixed interval data based equations, a form of Schumacher model was tried with both data sets for net basal area/ha and mean top height equations as shown in 9.1 and 9.2.

$$G_2=G_1(T1/T2)^{\beta 1}e(\alpha+ALT\beta_2+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(T_1/T_2)^{\beta 1}) \quad (9.1)$$

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1)(1-(T_1/T_2)^{\beta}) \quad (9.2)$$

where G, h_{100} and ALT are basal area/ha, mean top height and altitude respectively, while K1, K2 and K3 are dummy variables for location. The confidence intervals for both models showed that all estimates are significant at α=0.05 level, as none included zero, but the 95 percents confidence intervals for the mixed interval model

are much tighter. The residual patterns of both two models were good and without bias.

Long interval data sets, containing measurement pairs all with intervals more than 8 years, were utilized to compare long term projection capabilities with both the short interval and mixed interval based models. The residual patterns for the mixed interval data equation was much tighter than that of the short equation, when they were fitted to the long interval data. The mean residual value for the short set equation was also greater than for the mixed set equation for the net basal as well as the mean top height equation. Skewness for the mixed set was, moreover, much closer to a 0 value as shown in Table 9.1.

Table 9.1 Mean residual and skewness for short and mixed data sets equations after being applied to the long data set

Variables	Net basal area/ha		Mean top height	
	Short data equation	Mixed data equation	Short data equation	Mixed data equation
Mean residuals	0.6905	0.6159	0.4385	0.0414
Skewness	-0.2082	0.1121	0.4089	0.2608

9.1.1 Net Basal Area Equation

Several other equations were further analysed in projection form for predicting net basal area/ha. A modified Schumacher polymorphic function 9.3 for the short and Gompertz polymorphic function 9.4 for the long and mixed interval were found to give the best fits for net basal area growth. The modified Schumacher was of the form:

$$G_2=G_1(T_1/T_2)^{\beta 1}\exp((\alpha+ALT\beta_2+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(T_1/T_2)^{\beta 1})+\beta_6 Xt(1/T_2-1/T_1)Tt/T_2) \quad (9.3)$$

where the term Xt represents a thinning index not included in equation 9.1.

The Gompertz functional form was

$$G_2 = G_1 \, e^{(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2))} \, e^{(\alpha + \beta_4 ALT + \beta_1 K_1 + \beta_2 K_2 + \beta_3 K_3)(1 - \exp(-\beta (T_2 - T_1) + \gamma (T_2^2 - T_1^2)))} \quad (9.4)$$

where G, T and ALT are basal area/ha, stand age and altitude respectively, K1, K2 and K3 are dummy variables accounting for location and Xt is a thinning index, defined in this study as the ratio of quadratic mean diameter of trees removed in thinning to the quadratic mean diameter of the stand before thinning. When the thinning effect was examined through a thinning index, its coefficient was not found to be all that important for the long interval data set.

When site index was introduced to equation 9.3, Temu's apparently anomalous result from his 1992 research was confirmed. Altitude above sea level was introduced because Douglas-fir basal area growth is sometimes greater at higher altitudes, but higher altitudes are confounded by locality, which reflects varying rainfall gradients.

There were improvements of around 14 and 11 percent in the sums of squares for the long and mixed interval data sets equations respectively after including the dummy variables, which allowed, therefore, different asymptotes for basal area/ha regionally (K1, K2 and K3 for Nelson, Southland and Westland respectively).

The basal area/ha projection equation includes thinning as a component of the thinning term with the short interval data set. Development of the basal area/ha equation depends on the kind of thinning and the time elapsed since the last thinning occurred. The thinning term was not helpful in the model fitted with long interval data because the effects of thinning on productivity are short-lived and it is therefore difficult to detect thinning effects with long measurement intervals.

Altitude, thinning index and dummy variables were also found to improve the model in the mixed interval data set, when they were introduced independently to the basic polymorphic Gompertz equation, however, were not superior to the equation which included only the three dummy variables and altitude.

When equations 9.3 and 9.4 were compared to Temu's and Law's overall model they were found to be more precise and less biased for all three interval data sets than (see Table 7.1).

9.1.2 Mean Top Height Equation

A modified Schumacher polymorphic function 9.5 gave the best fit for the short interval data set:

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+\beta_2 ALT+\beta_1 K_1)(1-(T_1/T_2)^{\beta}) \qquad (9.5)$$

while the same functional form, but with two additional dummy variables and without altitude for the long and mixed interval data sets proved best for them:

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+K_3\beta_3)(1-(T_1/T_2)^{\beta}) \qquad (9.6)$$

where h_{100}, ALT and T are mean top height, altitude and stand age respectively and K1, K2 and K3 are dummy variables which allowed for different asymptotes for mean top height regionally. Only one dummy variable for the Nelson region was necessary in the short interval data set equation as shown in equation 9.5, but there were significant differences between coefficients for each of the Canterbury, Nelson, Southland and Westland regions for the long and mixed interval data sets.

Altitude, which Temu had found to be an important variable for explaining variation in mean top height growth, was not introduced to the long and the mixed sets equations in the final formulation, because it was found to be unhelpful when used together with dummy variables.

Comparison of the equations for each interval data set with the existing Temu and Law models, confirmed that they were also more precise than the corresponding Temu and Law models as shown in Table 7.1.

9.1.3 Mortality Equation

A modified form of exponential equation 9.7 gave the best fit for projecting survival of stems/ha in the short data set:

$$N_2 = N_1 \exp((\alpha + \gamma K1)(T_2^{\beta} - T_1^{\beta})) \quad (9.7)$$

For the long and mixed interval data sets, a modified Gompertz, equations 9.8 which including site index and dummy variables was fitted:

$$N_2 = N_1 e^{(\beta(T_2^2 - T_1^2))} e^{(\alpha+\beta_1 K_1+\beta_2 K_2+\beta_3 K_3+\beta_4 S)(1-\exp(\beta(T_2^2 - T_1^2)))} \quad (9.8)$$

There were significant differences in mortality trends among regions and the values of asymptotes were clearly related to site indices in the long and mixed interval equations. A dummy variable was necessary only for the Nelson region in the short interval equation.

Comparison of the equations for each interval data set with the existing Temu and Law models revealed that the new models were more precise than Law's model for each data set and achieved better precision than Temu's model for the short and mixed interval data. Temu's model, however, gave at least slightly better precision than the best model for the long interval data set.

9.2 Crossing Fitting

In order to examine consistency among the predicting equations which had been developed for the various interval data sets, each equation was applied to the other data sets.

9.2.1 Net Basal Area equation

When the best short interval basal area/ha equation, 9.3, was applied to the long and mixed interval data, the residual patterns were biased, and 95 percent residuals lay within mainly ±13.03 m^2/ha and ±9.89 m^2/ha respectively. Corresponding values for the best long and mixed interval models were ±11.36 m^2/ha and ±7.95 m^2/ha.

When the basal area equation 9.4, which proved best for the long interval data set, was applied to the short interval data, it showed generally a good fit and resulted in unbiased residuals pattern. Applied to the mixed interval data set, it resulted in a bit similar to that of the best model specifically for the mixed interval data set, without a biased residual pattern.

Equation 9.4 also best represented for the mixed interval data basal area/ha equation. When this equation applied to the short interval data set, the residuals pattern were well balanced and without bias. It gave also a good residual pattern when applied to the long interval data (see Figures 5.5 and 5.6). Applying this equation to the short and long interval data sets, 95 percent residuals lay within ±2.43 m^2/ha and 11.45 m^2/ha respectively, while corresponding residuals for the best models in each data set were ±1.92 and 11.36 m^2/ha

Applying the mixed best model to the short and long data sets showed generally good fits without bias, but applying the best model in the short data set to the other data sets showed bias. In considering long term applicability between the short and mixed models, it was evident showed that the mixed model gave a better fit than the short data model, as shown in the residuals pattern and in Table 9.2.

Table 9.2 Summary of applying the best mixed and the best short to long data set

Variable	Short data model	Mixed data model
Mean residual error (m^2/ha)	0.66809	0.01845
Highest negative residual (m^2/ha)	-29.5	-21.7
Highest positive Residual (m^2/ha)	26.6	21.7
Skewness	-0.3	0.06
Kurtosis	0.89	0.32

It can, therefore, be concluded that by using the mixed data set, which contains a wide range of (T_2- T_1) time intervals, the precision of estimating parameters could be increased and an overall more precise projection equation derived.

9.2.2 Mean Top Height Equation

The best mean top height equations for each interval data sets were of a Schumacher polymorphic functional form. The only difference was that two additional dummy variables were needed for the long and mixed interval data set. Altitude was found to be an important explanatory variable for the short interval data. Because of the similar functional forms, cross-fitting among the different data sets for mean top height resulted in no abnormality in residuals pattern and showed consistency over the predicting equations which were developed for all three interval data sets.

The mixed interval mean top height equation produced nearly the same values of mean residual error and skewness as those for the short data best model when applied to its data. The short interval model did not produce as good a result for the mixed interval data set as shown Table 5.4. The mixed interval model, thus, proved superior to the equations derived from other kinds of data sets for predicting mean top height.

9.2.3 Stem Survival/ha Equation

When the best short interval stem survival/ha equation, 9.7 was applied to the other data sets, the residuals patterns were poor and largely biased, as shown in Figures 5.13 and 5.14. The precision was also poorer, as this equation had a mean residual error of ± 31 stems/ha for the long interval and ± 21 stems/ha for the mixed interval data, while the best model for the long and mixed interval had ± 5 stems/ha and ± 1 stem/ha respectively.

A modified Gompertz equation, 9.8, represented the best stem survival/ha equation for the long and mixed interval data sets. When this equation was applied to the short interval data set, it fitted well without apparent bias as, shown in Figure 5.15. Ninety five percent of residuals lay within ± 63 stems/ha, compared with the corresponding best short interval equation in which 95 percent residuals lay within ± 55 stems/ha

When equation 9.8 was applied to the long and mixed interval data with their own coefficients, the residuals patterns were nearly the same as those for the model chosen as best for their own data sets because of the same functional forms and similar coefficients (see Tables 4.19 and 4.28)

9.3 Growth and Yield Model with a Data Set Free from Auto-Correlation

A subset of uncorrelated data was created through selecting one remeasurement from each PSP which was then used to validate the appropriateness of each of the above equations, which had been derived from full, but correlated data.

9.3.1 Net Basal Area/ha Equation

A modified Schumacher polymorphic function 9.9 was found to give the best fit to the data set free from auto-correlation.

$$G_2=G_1(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+k_3\beta_3)(1-(T_1/T_2)^{\beta}) \quad (9.9)$$

Altitude improved the mean square error (MSE) value when it was introduced to the basic Schumacher equation, but it was not significant when used together with dummy variables. The thinning index and site index were also insignificant at the α=0.05 level. The equation for this data set free from auto-correlation had minimum residuals of –23.05 m^2/ha and 20.12 m^2 /ha, while 95 percent of residuals lay within ±8.72 m^2/ha.

When equation 9.9 was fitted with short, long, and mixed interval data sets which had the same number of observations as those in the data set free from auto-correlation (i.e. the same degrees of freedom), the confidence intervals for all models showed that all estimates were significant at α=0.05 level. The 95 percent confidence interval for the data set free from auto-correlation equation had almost the same tightness as that of the short, long and mixed data sets. The data set free from auto-correlation equation was slightly tighter than the short interval equation, and slightly looser than the long and mixed interval equations (see Table 6.5).

9.3.2 Mean Top Height Equation

A modified Schumacher polymorphic equation 9.10 was found to give the best fit for the data set free from auto-correlation.

$$h_{100,2}=h_{100,1}(T_1/T_2)^{\beta}e(\alpha+K_1\beta_1+K_2\beta_2+k_3\beta_3)(1-(T_1/T_2)^{\beta}) \quad (9.10)$$

This equation is of the same form as the best in the long and mixed interval data sets for mean top height and is of a form similar to the best in the short interval data set which has, however, only one dummy variable representing locality.

Equation 9.10 was fitted to two data sets: one contained the same number of observations in each data set; the other had the same number of observations as those in the data set free from auto-correlation in order to compare the coefficients of

equations thoroughly. It was clear that they had similar regression coefficients, as shown in Tables 6.8 and 6.9. The significance levels of the coefficients of the mean top height equation derived from the data free from auto-correlation, was compared with short, long and mixed interval data equations using the same degrees of freedom. The 95 percent confidence interval for the data free from auto-correlation equation showed slightly more tightness than the short interval equation, while the long and mixed interval equations were slightly tighter than the data free from auto-correlation equation. The equation for the data set free from auto-correlation had minimum and maximum residuals of -5.17 m and 4.73 m respectively, while 95 percent of residuals were contained within ± 2.26 m.

9.3.3 Mortality Equation

Equation 9.11 gave the best fit for stem survival/ha for the data set free from auto-correlation.

$$N_2=N_1e^{((\alpha+\beta_1K_1+\beta_2K_2+\beta_3K_3)(T_2-T_1))} \qquad (9.11)$$

There were differences in mortality trends among the Canterbury, Nelson, Southland and Westland regions. The equation had minimum and maximum residuals of -578 stems/ha and 515 stems/ha, while 95 percent of residuals were contained ±316 stems/ha. It was not possible to compare the coefficients of this equation with those of the short, long and mixed interval data sets because the best model in the each data set had different forms of equation. Equation 9.12, therefore, which was used for comparative purposes, was able to fit all four data sets satisfactorily even though it was not the best model in each data set. When the coefficients of equations were compared they were found to be similar coefficients (see Table 6. 12).

$$N_2=N_1e^{(\alpha(T_2-T_1))} \qquad (9.12)$$

Conclusions

The revised models derived in this study produced accurate and precise stand level yield forecasts for each of the different interval data sets for Douglas-fir stands throughout the South Island of New Zealand. Findings from comparing models obtained from different data sets are outlined below.

1. It was possible to derive a more precise and robust model than Temu's and Law's models by forming three data sets, which represented different interval lengths between the re-measured growth data. The model recommended as best here is that derived from the mixed intervals which contains a range of T_2- T_1 time intervals ranging from 1 to 28 years.

2. When the length of the projection interval increased, precision for the whole stand growth model apparently decreased but without a corresponding decline in accuracy.

3. Dummy variables representing region proved important predictor variables in the stand growth models for estimating basal area/ha, mean top height and stem survival.

4. When applying the best short interval equation for basal area/ha and survival stems/ha to the other data sets, residual patterns were poor and biased. But when the best long and mixed interval equations were used for the short interval data set the residual patterns were unbiased and showed a good fit. Thus, the mixed data model, which has more precise parameters than for the long interval set, proved to be more applicable to the short data set.

5. The best equation fits for basal area/ha, mean top height and stem survival/ha in the long and mixed interval data sets were of the same functional form and had almost the same parameters. When these equations were applied to the short

interval data set, they presented the same overall result and fitted well without apparent bias. When these equations were fitted crosswise with their own coefficients, the result was nearly same as for the best fitting model in each data set.

6. When the independent data set with its less correlated plot data was used to test the effects of eliminating auto-correlation, it produced similar regression coefficients to the full data set, though the precision was, of course, much poorer. But this part of the analysis confirmed the robustness of the findings pertaining to the full data sets.

7. Predictor variables affected the asymptotes of equations rather than the shapes of growth trajectories. This trend was obvious in the long and mixed interval data sets but it was not so distinct for the net basal area equation derived from the short interval data set.

The use of a mixed interval data set, which contained a range of measurement periods, from the short to long term, which has been found to increase precision in estimating model coefficients without incurring bias is thus advocated for future growth and yield studies.

Acknowledgements

First of all, I would like to thank the Almighty God who granted me the courage and knowledge to complete this study and the desire to overcome the moral and physical difficulties of undertaking research in a foreign land.

I wish to thank my supervisor Dr. Euan G. Mason for his constructive encouragement, advice, comments and continued support throughout this study.

I would like to express my sincere thanks to Dr. A. Graham D. Whyte for his continued guidance and positive support, and especially for his tireless efforts during the writing and comprehension of my thesis.

Many thanks to the New Zealand Forest Research Institute for providing me with the Permanent Sample Plots data that was a vital part of my research. Also earnest thanks to the following New Zealand private forestry companies for permitting me the use of data concerning South Island Douglas-fir plantations: Carter Harvey Ltd.; Ernslaw One Ltd.; Fletcher Challenge Forests Ltd.; Rayonier N.Z Ltd.; Resource Management Ltd.; Timberlands West Coast Ltd.; and Wenita Forest Products Ltd..

I deeply appreciated the friendly attitude and moral support of Prof. R. Sands. I also thank Victoria Mackisack for her administrative assistance as well as her friendly welcome and support for my family.

Special thanks to the all staff of the School of Forestry, University of Canterbury, for their friendship and assistance.

Many thanks are also due to all the forestry postgraduates and colleagues who I have been associated with during the past 4 years. I thank Balozi, Nishantha and Craig for their personal support throughout my study period.

I am highly grateful to my parents and parent in laws for their love, support and prayers while I have been in New Zealand. Also sincere thanks to my brothers, sisters, and many in-laws for their moral support.

Finally, I am thankful to my beloved wife Hyun Ju, and daughter Hye Won, for their love and patience. The sacrifice they have made cannot be expressed by recording a few words here, and their love is invaluable.

References

Adlard, P.G., 1995. Myth and reality in growth estimation. For. Ecol. Mgmt. 71:171-176.

Alder, D., 1980. Forest volume estimation and yield prediction. Vol. 2 – Yield prediction. FOA Forestry Paper 22.

Arney, J.D., 1974. An individual tree model for stand simulation in Douglas-fir. In: J. Fries(Ed). Growth Models for Tree and Stand Simulation. Royal College of Forestry, Stockholm, Sweden.

Assmann, E., 1970. The principles of forest yield study, Translated by S.H. Gardiner, Pargamon Press, Oxford; 506 pp.

Avery, T.E., and Burkhart, H.E., 1994. Forest measurement. McGraw-Hill Book Co., New York. 408 pp.

Bailey, R.L., and Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20: 155-159.

Bailey, R.L., and Dell T.R., 1973. Quantifying diameter distributions with the Weibull function. For. Sci. 19:97-104.

Bailey, R.L., Burgan, T.M., and Jokda, E.J., 1989. Fertilized mid-rotation aged slash pine plantations-stand structure and yield prediction models. Southern J. Appl. For. 13: 76-80.

Bailey, R.L., and Ware, K.D., 1983. Compatible basal area growth and yield model for thinned and unthinned stands. Can. J. For. Res. 13: 563-571.

Barrett, J.W., 1980. Regional Silviculture of the United States. John Wiley and Sons, Toronto. 551 pp.

Beck, D.E., 1971. Polymorphic site index curves for white pine in the southern applications. USDA, For. Serv. Res. Pap. Se-80.

Beekhuis, J., 1978. Growth decline in Douglas fir. NZFRI symposium No. 15.

Belcher, E.M., Holdaway, M.R., and Brand, G.J., 1982. A description of STEMS: the stand tree evaluation and modelling system. USDA, For. Serv. Gen. Tech. Rep. NC-79; 18p.

Belli, K.l., and Nautiyal, J.C., 1989. Production functions: a useful tool for forest management. For. Ecol. Mgmt 29: 267-275.

Bennet, F.A., McGee, C.E., and Clutter, J.L., 1959: Yield of old – field slash pine plantations. USDA, For. Serv. S.E. Exp. Stn. Res. Pap. 107: 19 p.

Berkey, C.S., 1982. Bayesian approach for non-linear growth model. Biometrics 38: 953-961.

Bickford, C.A., Baker, F.S., and Wilson, F.G., 1957. Stocking, normality, and measurement of stand density. J. For. 55: 99-104.

Borders, B.E., Bailey, R.L., and Ware, K.D., 1984. Slash pine index from a polymorphic model joining (splining) non-polynominal segments with an algebraic difference method. For. Sci. 30(2): 411-423.

Bossel, H., 1991. Modelling forest dynamics: Moving from description to explanation. For. Ecol. Mgmt. 42: 129-142.

Bowling, E.H., Burkhart, H.E., Burk, T.E., and Beck, D.E., 1989. A stand-level multispecies growth model for application hardwoods. Can. J. For. Res. 19 (4): 405-412.

Brickell, J.E., 1968. A method of constructing site index curves from measurements of tree age and height – its application to inland Douglas-fir. USDA, For. Serv. Ge. Tech. Rep. PNW. 57: 27 p.

Bruce, D., and Wensel, L.G., 1987. Modelling forest growth: Approaches, definitions, and problems, In:Ek, A.R., Shifley, S.R. and Burk, T.E. (Eds.), Proceedings of the IUFRO symposium on Forest growth modelling and prediction, Minneapolis, Minnesota, USDA, For. Serv. Gen. Tech. Rep. NC-120: 1-8.

Buckman, R.E., 1962. Growth and yield of red pine in Minnesota. USDA, For. Serv. Tech. Bull. 1272.

Buckmam, R.G., and Shifley, S.R., 1983. Guide to evaluating growth projection systems. J. For. 81(4): 232-234.

Burkhart, H.E., and Tennent, R.B., 1977. Site index equations for radiata pine in New Zealand. N.Z. J. For. Sci. 7(3): 408-416.

Burns, R.M., 1983. Silvicultural systems for the major forest types of the United States. Agriculture Handbook No. 445, USDA, Washington, D C. 191 pp.

Carmean, W.H., 1972. Site index curves for upland oaks in the central states. For. Sci. 18: 109-120.

Causton, D.R., 1983. A biologist's basic mathematics. Edward, Arnold, Balimore, Md., USA. 216 pp.

Chikono, C., 1994. Modelling the growth of mean top height and basal area of Eucalyptus gradis in Zimbabwe. Msc. Report, University of Canterbury, New Zealand. 66p.

Clutter, J.L., 1963. Compatible growth and yield modes for loblolly pine. For. Sci. 9: 354-371.

Clutter, J.L., and Allison, B.J., 1974. A growth and yield model for Pinus radiata in New Zealand. In: J. Fries(Ed). Growth Models for Tree and Stand Simulation. Royal College of Forestry, Stockholm, Sweden. 136-160.

Clutter, J.L., and Bennet, F.A., 1965. Diameter distribution in old-slash pine plantations. Ga. For. Res. Counc. Rep. 23. 9 p.

Clutter, J.L., and Jones, E.P., 1980. Prediction of growth after thinning in old-field slash pine plantations. USDA, For. Serv. Res. Pap. SE-217.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L., 1983. Timber management: a quantitative approach. John Wiley and Sons, New York. 333 pp.

Cocharn, W.G., 1977. Sampling techniques. John and Wiley and Sons, New York. 428 pp.

Coile, T.S., 1948. Relation of soil characteristics to site index of loblolly and shortleaf pines in the lower piedmont region of North Carolina. Duke Univ. School of For. Bull. 13.

Cown, D.J., 1992. New Zealand radiata pine and Douglas-fir: Suitability for processing. N.Z Ministry of Forestry, Forest Research Institute. FRI Bull. No. 168.

Curtis, R.O., 1972. Yield tables past and present. J. For. 70 (1): 28-32.

Daniels, R.F., and Burkhart, H.E., 1988. An integrated system of forest stand models. For. Ecol. Mgmt. 23: 159-177.

Davis, A.W. and West, P.W., 1981. Remarks on " Generalised least-squares estimation of yield functions" by I.S. Ferguson and J.W. Leech. For. Sci. 27: 233-239.

Doolittle, W.T., 1958. Site index comparisons for several forest species in the southern Appalachians. Proc. Soil Sci. Soc. Amer. 22:455-458.

Draper, N.R. and Smith, H., 1981. Applied regression analysis (second edition), John Wiley and Sons, New York. 709 pp.

Duff, G., 1956. Yield of unthinned Douglas fir, Corsican pine and Ponderosa pine in New Zealand. N.Z. Forest. Service. Note No. 5.

Elliott, D.A., 1969. A provision variable density yield table for Douglas fir in Kaingaroa. N.Z. J. For. 14(1): 75-86.

Ferguson, I.S., 1979. Growth functions for radiata pine plantations. In H.L. Wright(Ed) Proceedings IUFRO S4.01, Oxford, Sept. 1979:25-45.

Ferguson, I.S. and Leech, J.W., 1978. Generalised least-squares estimation of yield functions. For. Sci. 24: 27-42.

Garcia, O., 1979. Modelling stand development with stochastic differential equations. In Elliot, D.A.(Ed). Mensuration for management planning of exotic forest plantations. N.Z. For. Serv., FRI Symposium. No. 20: 315-333.

Garcia, O., 1984. New class of growth models for even-aged stands – pinus radiata in Golden Downs Forest. N.Z. J. For. Sci. 14(1): 65-88.

Garcia, O., 1987. Experience with an advanced growth modelling methodology. In:Ek, A.R., Shifley, S.R. and Burk, T.E. (Eds.), Proceedings of the IUFRO symposium on Forest growth modelling and prediction, Minneapolis, Minnesota, USDA, For. Serv. Gen. Tech. Rep. NC-120: 668-675.

Garcia, O., 1988. Growth modelling – a (re)view, N. Z. Forestry 33 (3): 14-17.

Garcia, O., 1991. What is a diameter distribution?, In: Proceedings of the IUFRO symposium on Integrated Forest management Information System, Tsukuba, Japan: 11-29.

Garcia, O., 1994. The state-space approach in growth modelling.Can. J. For. Res. 24: 1894-1903.

Gertner, G.Z., 1984. Localising a diameter increment model with a sequential Bayesian procedure. For. Sci. 30(4): 851-864.

Goulding, C.J., 1979. Validation of growth models for pinus radiata in New Zealand. N.Z. J. For. 24(1): 108-124.

Goulding, C.J., 1994. Development growth models for pinus radiata in New Zealand experience with management and process models. For. Ecol. Mgmt. 69. 1-3: 331-343.

Goulding, C.J., and Shirey, J.W., 1979. A method to predict the yield of log assorments for long term planning. In Elliott, D.A.(Ed). Mensuration of management planning of exotic forest plantations. N.Z. For. Serv., FRI Symposium No. 20: 301-314.

Green, E.J. Strawdermann, W.E., and Thomas, C.E., 1992. Empirical Bayes development of Honduran pine yield models. For. Sci. 38(1): 21-33.

Gujarat, D., 1970. Use of dummy variables in testing for equality between sets of coefficients in liner regression: a generalization. American statistician 25(4) 21-33.

Hafley, W.L., and Schreuder, H.T., 1977. Statistical distribution for fitting diameter and height data in even aged stands. Can. J. For. Res. 7:481- 487.

Hall, C.A.S., and Day, Jr. J. W., 1977. Systems and models: terms and basic principles. In Charles A.S. Hall and John. W. Day Jr. Ecosystem modelling, in theory and practice: an introduction with case histories. University Press of Colorado. 684 pp.

Hegyi, F., 1974. A simulation model for managing jack-pine stands. In: J.Fries(Ed). Growth Models for Tree and Stand Simulation. Royal College of Forestry, Stockholm, Sweden.

Hellawell, C.R., 1978. Douglas fir as a sources of timber for engineering purpose. NZFRI symposium No. 15: 240-252.

Hood, I.A., and Kershaw, K.J., 1975. Distribution and infection period of Phaeocryptopus gaeumannii in New Zealand. For. Sci. 5(2): 201-208.

Hosmer, D.W., and Lemeshow, S., 1989. Applied logistic regression. John Wiley and Sons, Inc New York. 307 pp.

Hunter, I.R., and Gibson, A.R., 1984. Predicting pinus radiata site index variables from environmental variables, N.Z. J. For. Sci. 14(1): 53-64.

Husch, B., 1971. Planning a forestry inventory. FAO, UN, Rome. 121 pp.

Hyink, D.M., and Moser Jr, J.W., 1983. A generalised framework for projecting forest yield and stand structure using diameter distributions. For. Sci. 29: 85-99.

James, R.N., 1975. A review of Douglas-fir in new Zealand. N.Z. J. For. 20(1): 107-128.

James, R.N., and Bunn, E.H., 1978. Editors summary. NZFRI symposium No. 15.

Knowles, R.L., McInnes, I.P., and Miller, J.T., 1995. Silviculture of Douglas-fir. In: Hammond, D, (Ed). Forestry Handbook. New Zealand Institute Forestry. 87-89.

Kuru, G.A., 1989. A diameter distribution growth and yield model for naturally regenerated and seeded stands of radiata pine in Kaingaroa forest. Msc. Thesis, School of Forestry, University of Canterbury, New Zealand. 102 pp.

Kuru, G.A., Whyte, A.G.D., and Woollons, R.C., 1992. Utility of reverse Weibull and extreme value density functions to refine diameter distribution growth estimates, For.Ecol. Mgmt. 48. 1-2, 165-174.

Landsberg, J.J., 1986. Physiological ecology of forest production. Academic press. London. 165 pp.

Lanner, R.M., 1985. On the intensity of height growth to spacing. For. Ecol. Mgmt. 13:143-148.

Law, K.R.N., 1990. A growth model for Douglas fir grown in the South Island of New Zealand. FRI/Industry Research Cooperatives: Stand Growth Modelling Cooperative Report No. 18. Forest Research Institute, Private Bag, Rotorua.

Leary, R.A., 1979. Design. In: A generalized forest projection system allied to the Lake States region. USDA, For. Serv. Gen. Tech. Rep. NC-49: 5-15.

Lemmon, P.E., and Schumacher, F.X., 1962: Volume and diameter growth of Ponderosa pine trees as influenced by site index, density, age and size. For. Sci. 8:236-249.

Lin, J.Y., 1974. Stand growth simulation models for Douglas-fir and Western hemlock in the Northwestern United states. In: J.Fries(Ed). Growth Models for Tree and Stand Simulation. Royal College of Forestry, Stockholm, Sweden.

Liu xu, Wood, G.R., Woollons, R.C., and Whyte, A.G.D., 1992. Stand table prediction with reverse Weibull and extreme value density functions: Some theoretical considerations. For. Ecol. Mgmt. 48. 1-2: 175-178.

Liu, Xu., 1990. Growth and yield of Douglas fir plantations in the Central North Island of New Zealand. Ph.D Thesis. School of Forestry, University of Canterbury, New Zealand. 244 pp.

MacKinney, A.L., Schumacher, F.X., and Chaiken, L.E., 1937. Construction of yield tables for nonnormal loblolly pine stands. Jour. Agr. Res. 54; 531-545.

MacKinney, A.L., and Chaiken, L.E., 1939. Volume, yield and growth of loblolly pine in the mid – Atlantic coastal region. U.S. For. Serv. App. For. Exp. Stn. Tech. Note 33: 30 pp.

Mason, E.G., 1992. Decision support systems for establishing radiata pine in Central North Island of New Zealand. Ph.D Thesis. School of Forestry, University of Canterbury, New Zealand. 301 pp.

Matney, T.G., and Sullivan, A.D., 1982. Compatible stand and stock tables for thinned loblolly pine stands. For. Sci. 28: 161– 71.

McArdle, R.E., Meyer, W.H. and Bruce, D., 1961. The yield of Douglas-fir in the Pacific Northwest. USDA Tech. Bull. No. 201.

McComb, W.C., Spies, T.A., and Emmingham, W.H., 1993. Douglas-fir forests: Managing for timber and mature-forest habitat. J. For. 91:31-42.

Miller, J.T., and Knowles, F.B., 1994. Introduced forest trees in New Zealand: Recognition, Role, and Seed Sources, Douglas-fir Pseudotsuga menziesii (Mirbel) Franco. NZFRI Bull. No.168.

Monserud, R.A., 1985. Comparison of Douglas fir site index and height growth curves in Northwest. Can. J. For. Res. 15(4): 673-679.

Monserud, R.A., 1987. Variations on a theme of site index. USDA, For. Serv. Tech. Rep. NC-120: 419-427.

Morhren, G.M.J., and Burkhart, H.E., 1994. Contrasts between biologically-based models and management-oriented growth and yield models. For. Ecol. Mgmt. 69. 1-3:1-5.

Mountfort, C.J., 1978. Growth of Douglas fir. N.Z. For. Serv. FRI. Symposium No. 15: 375-411.

Munro, D.D. 1974. Forest growth models – a prognosis. In: J.Fries(Ed). Growth Models for Tree and Stand Simulation. Royal College of Forestry, Stockholm, Sweden.

Murphy, P.A., and Farrar, Jr., R.M., 1988. Basal area projection equations for thinned natural even-aged forest stands. Can, J. For. Res. 18: 827-832.

Nelson, T.C., 1964. Diameter distribution and growth loblolly pine. For. Sci. 10: 105-115.

Neilson, D.A., and Business media services limited., 1994. The New Zealand forestry sector: an international prototype for a sustained resource industry. D.A. Neilson and Associate/BMS Ltd. Rotorua, New Zealand.

Newberry, J.D., and Stage, A.R., 1987. Validating forest growth models. Procedures defined by Resources decisions. In:Ek, A.R., Shifley, S.R. and Burk, T.E. (Eds.), Proceedings of the IUFRO symposium on Forest growth modelling and prediction, Minneapolis, Minnesota, USDA, For. Serv. Gen. Tech. Rep. NC-120: 786-793.

Ngugi, M.R., 1996. Growth and yield models for Cupress lusitanica (Mill.) and Pinus patula (Schlect & Cham.) grown in Kenya. Msc. Thesis. School of Forestry, University of Canterbury, New Zealand. 142 pp.

Nokoe, S., 1978. Demonstrating the flexibility of the Gompertz function as a yield model using mature species data. Comm. For. Rev. 57:35-42.

Nystrom, K., and Kexi, M., 1997. Individual tree basal area growth models for young stands of Norway spruce in Sweden. For. Ecol. Mgmt. 97: 173-185.

Olderwald, R.G., and Hans, R.P., 1993. Corroborating models with model properties. For. Ecol. Mgmt. 62: 271-283.

Oliver, C.D., Honley, D.P., and Johnson, J.A., 1986. Douglas-fir: stand management for the future., 1986. Instotute of Forest Resources Contribution No. 55. College of Resources, University of Washington.

Philip, M.S., 1994. Measuring Trees and Forests 2nd ed. CAB International, UK. 310 pp.

Pienaar, L.V., 1979. An approximation of basal area growth after thinning based on growth in unthinned plantations. For. Sci. 25: 223-232.

Pienaar, L.V., 1989. A stand table projection approach to yield prediction in plantations. South African Forestry. Journal. 149: 44-47.

Pienaar, L.V., and Turnbull, K.J., 1973. The Chapman-Richards generalization of von Bertalanffy's model for basal area growth and yield in even-aged stands. For. Sci. 19: 2-22.

Pienaar, L.V., and Harrison, W.M., 1988. A stand table projection approach to yield projection in unthinned even-aged stands. For. Sci. 34(3): 804-808.

Pienaar, L.V., Shiver, B.D., and Grider, G.E., 1985. Predicting basal area growth in thinned slash pine plantations. For. Sci. 31(3): 731-741.

Piennar, L.V., and Shiver, B.D., 1984. An analysis of basal area growth in 45-year-old unthinned and thinned slash pine plantation plots. For. Sci. 30: 933-942.

Ralston, M.L., and Jenrich, R.I., 1979. Dud. A derivative-free algorithm for nonlinear least-squares. Technometrics. 20(1): 7-14.

Ramirez-Maldonado, Hugo, Bailey, R. L., and Borders, B.E., 1987. Some implications of the algebraic difference approach for developing growth models. In:Ek, A.R., Shifley, S.R. and Burk, T.E. (Eds.), Proceedings of the IUFRO symposium on

Forest growth modelling and prediction, Minneapolis, Minnesota, USDA, For. Serv. Gen. Tech. Rep. NC-120: 731-738.

Reineke, L.H., 1933. Perfecting a stand density index for even–aged forests. J. Agric. Res. 46: 627-638.

Rennolls, K., 1995. Forest height growth modelling. For. Ecol. Mgmt. 71: 217-225.

SAS Institute Inc., 1990. SAS/STAT User's guide, Version 6, Cary, NC.

SAS Institute Inc., 1990. SAS/STAT Procedures guide, Version 6, NC.

Schumacher, F.X., 1939. A new growth curve and its application to timber-yield studies. J. For. 37: 819-820.

Schumacher, F.X., and Colie, T.S., 1960. Growth and yield of natural stands of the Southern pines. T.S. Colie, Inc., Durham, N. C. 115 p.

Shifley, S., and Fairweather, S., 1983. Individual-tree diameter growth and mortality models for western Oregon. In proceeding of renewable resource inventories for monitoring changes and trends. Corvallis, Oregon. 737 pp.

Smith, W.B., 1983. Adjusting STEMS regional forest growth model to improve local predictions. USDA, For. Serv. Res. Note. NC-197.

Spurr, S.H., 1961. Observation on Douglas fir in New Zealand. N.Z. For. Serv., Tech. Pap. No. 39.

Spurr, S.H., 1963. Growth of Douglas fir in New Zealand. N.Z. For. Serv. Tech. Pap. 43.

Spurr, S.H., and Barnes, B.V., 1980. Forest ecology. 3nd ed. Wiley, New York. 687 pp.

Stage, A.R., 1973: Proggnosis model for stand development. USDA For.Serv. Res. Pap. INT-137. 32p.

Sullivan, A.D., and Clutter, J.L., 1972. A simultaneous growth and yield model for loblolly pine. For. Sci. 18: 76-86.

Temu, M.J., 1992. Forecasting yield of Douglas fir in the South Island of New Zealand, Ph.D Thesis, School of Forestry, Canterbury University, New Zealand, 246 pp.

Tennent, R.B., 1982. Individual tree growth model for pinus radiata. N.Z. J. For. Sci. 12(1): 62-70.

Tennent, R.B., 1995. An examination of the validity of three Douglas fir growth models. FRI/Industry Research Cooperatives: Stand Growth Modelling Cooperative Report No. 41. Forest Research Institute, Private Bag, Rotorua.

Umemura, T., 1984. An entirely new growth curves based on a second order linear differential equation. Proc IUFRO Symp. For. Mang. Plan and Mang. Econ. Uni. Tokyo, Japan, Oct. 15-19. 1984: 591-597.

Ure, J., 1950. The natural vegetation of the Kaingaroa plains as an indicator of site quality for exotic conifers. N. Z. J. For. 6: 112-123.

Vanclay, J.K. and Skovsgaard, J.P., 1996. Evaluating forest growth models. Forest ecology and management: in press.

Vancly J.K., 1994. Modelling forest growth and yield: application to mixed tropical forests. CAB International Wallingford, UK. 312 pp.

Villanueva, T.R., 1992. Integrated yield forecasting and harvest scheduling in tropical pine plantation in Fiji, Ph.D. Thesis, School of Forestry, University of Canterbury, New Zealand. 208pp.

Villanueva, T.R., and Whyte, A.G.D., 1992. Integrated yield forecasting and harvest scheduling for tropical plantations, In: Whyte, A.G.D., (Ed.) proceedings of IUFRO symposium on " Integrated decision-making in planning and control of forest operations", University of Canterbury, New Zealand. 82-95.

Ware, K.D., Boarder, B.E., and Bailey, R.L., 1988. Estimating growth and yield of thinned slash pine plations on old field. In forest growth modelling and prediction. Vol.1 USDA, For. Serv. North Central For. Expt. Stn. Gen. Tech. Rep. NC-120: 33-36.

Whyte, A.G.D,, and Woollons, R.C. 1992., Diameter distribution growth and yield modelling: recent revisions and perspectives. School of forestry, University of Canterbury, New Zealand. Unpublished.

Whyte, A.G.D., Temu, M.J., and Woollons, R.C., 1992. Improving yield forecasting realiblity through aggregated modelling. In: G.B. wood, and B.J. Turner(Eds). proceedings IUFRO-Integrating Information Over space and Time; Australian National University Canberra, Jan;13-17, 1992:81-88.

Whyte, A.G.D., and Woollons, R.C., 1990. Modelling stand growth of radiata pine thinned to varying densities. Can, J. For. Res. 20: 1069-1076.

Wilcox, M.D., 1978. Genetic resources and supply of Douglas fir seed in New Zealand. NZFRI. Symposium No. 15: 210-217.

Woollons, R.C., 1988. Analysis and interpretation of forest fertilizer experiments. Ph.D. Thesis. School of Forestry, University of Canterbury, New Zealand. 256 pp.

Woollons, R.C., 1998. Evan-aged stand mortality estimation through a two-step regression process. For. Ecol. Mgmt. 105: 189-195.

Woollons, R.C., and Hayward, W.J., 1985. Revision of a growth and yield model for radiata pine in New Zealand. For. Ecol. Mgmt. 11:191-202.

Woollons, R. C., Snowdon, P. and Mitchell, N. D., 1998. Augmentating empirical stand projection equations with edaphic and climatic variables. For. Ecol. Mgmt. 98: (3) 267-275.

Woollons, R.C., and Wood, G.R., 1992. Utility and performance of five sigmoid yield-age functions fitted to stand growth data. Improving yield forecasting reliability through aggregated modelling. In: G.B. wood, and B.J. Turner(Eds). Proceedings IUFRO-Integrating Information Over space and Time; Australian National University Canberra, Jan;13-17, 1992: 71-80.

Woollons, R.C., Whyte, A.G.D., and Liu Xu., 1990. The Hossfeld function: an alternative model for depicting stand growth. Japanese Journal of Forestry 15: 25-35.

Yang, R.L., Kozak, A., and Smith, H.J.G., 1978. The potential of Weibull-type functions as flexible growth functions. Can. J. For. Res. 8:424-431.

Appendix I

Location and number of permanent Sample Plots

CANTERBURY	
Forest	Number of Permanent Sample Plots
Ashley	33
Geraldine	7
Hamner	14
Omihi	3
Total	57

NELSON	
Forest	Number of Permanent Sample Plots
Golden Downs	123
Ray Valley	10
Wairau	3
Total	136

SOUTHLAND	
Forest	Number of Permanent Sample Plots
Berwick	15
Blue Mount	52
Herbert	12
Hokonui	9
Longwood	28
Mount Allan	1
Naseby	4
Silverpeaks	7
Total	128

WESTLAND	
Forest	Number of Permanent Sample Plots
Granville	2
Hochstetter	13
Mawhera	12
Mahinapua	4
Victoria SFP	14
Total	45

Appendix II

Diskette – Files

The attached diskette consists of SAS files used to develop the equations and their outputs. Output files are stored with the extension *.LST. The names of these files and variables are described in chapter 3 and elsewhere in this thesis.

Appendix III

A stand simulation model for Douglas-fir grown in Canterbury, Nelson, Southland and Westland is presented with the file name NEWDFIR. EXE.

To run the model **Exit** from windows to **MS-DOS** and type **A:\NEWDFIR** and press the **ENTER** key. Input the required information to simulate stand outputs.